



Machine Learning Prediction of Cardiac Resynchronisation Therapy Response From Combination of Clinical and Model-Driven Data

Svyatoslav Khamzin¹, Arsenii Dokuchaev¹, Anastasia Bazhutina^{1,2}, Tatiana Chumarnaya¹, Stepan Zubarev³, Tamara Lyubimtseva³, Viktoria Lebedeva³, Dmitry Lebedev³, Viatcheslav Gurev⁴ and Olga Solovyova^{1,2*}

¹ Institute of Immunology and Physiology Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia, ² Ural Federal University, Yekaterinburg, Russia, ³ Almazov National Medical Research Centre, Saint Petersburg, Russia, ⁴ IBM Research, Yorktown, NY, United States

OPEN ACCESS

Edited by:

Hans Dierckx,
KU Leuven Kulak, Belgium

Reviewed by:

Steven Pogwizd,
University of Alabama at Birmingham,
United States
Jan Christoph,
University of California, San Francisco,
United States

*Correspondence:

Olga Solovyova
o.solovyova@iip.uran.ru

Specialty section:

This article was submitted to
Cardiac Electrophysiology,
a section of the journal
Frontiers in Physiology

Received: 04 August 2021

Accepted: 22 November 2021

Published: 14 December 2021

Citation:

Khamzin S, Dokuchaev A, Bazhutina A, Chumarnaya T, Zubarev S, Lyubimtseva T, Lebedeva V, Lebedev D, Gurev V and Solovyova O (2021) Machine Learning Prediction of Cardiac Resynchronisation Therapy Response From Combination of Clinical and Model-Driven Data. *Front. Physiol.* 12:753282. doi: 10.3389/fphys.2021.753282

Background: Up to 30–50% of chronic heart failure patients who underwent cardiac resynchronization therapy (CRT) do not respond to the treatment. Therefore, patient stratification for CRT and optimization of CRT device settings remain a challenge.

Objective: The main goal of our study is to develop a predictive model of CRT outcome using a combination of clinical data recorded in patients before CRT and simulations of the response to biventricular (BiV) pacing in personalized computational models of the cardiac electrophysiology.

Materials and Methods: Retrospective data from 57 patients who underwent CRT device implantation was utilized. Positive response to CRT was defined by a 10% increase in the left ventricular ejection fraction in a year after implantation. For each patient, an anatomical model of the heart and torso was reconstructed from MRI and CT images and tailored to ECG recorded in the participant. The models were used to compute ventricular activation time, ECG duration and electrical dyssynchrony indices during intrinsic rhythm and BiV pacing from the sites of implanted leads. For building a predictive model of CRT response, we used clinical data recorded before CRT device implantation together with model-derived biomarkers of ventricular excitation in the left bundle branch block mode of activation and under BiV stimulation. Several Machine Learning (ML) classifiers and feature selection algorithms were tested on the hybrid dataset, and the quality of predictors was assessed using the area under receiver operating curve (ROC AUC). The classifiers on the hybrid data were compared with ML models built on clinical data only.

Results: The best ML classifier utilizing a hybrid set of clinical and model-driven data demonstrated ROC AUC of 0.82, an accuracy of 0.82, sensitivity of 0.85, and specificity of 0.78, improving quality over that of ML predictors built on clinical data from much larger datasets by more than 0.1. Distance from the LV pacing site to the post-infarction

zone and ventricular activation characteristics under BiV pacing were shown as the most relevant model-driven features for CRT response classification.

Conclusion: Our results suggest that combination of clinical and model-driven data increases the accuracy of classification models for CRT outcomes.

Keywords: cardiac resynchronization therapy, machine learning, cardiac modeling, electrophysiology, prediction, heart failure, hybrid approach

1. INTRODUCTION

Cardiac resynchronization therapy (CRT) is one of the most effective non-pharmacological therapies for patients with chronic heart failure (CHF). It enhances the pumping function increasing the left-ventricular (LV) ejection fraction (EF), promotes reversed cardiac remodeling, and improves patients' quality of life (Abraham et al., 2002; Bristow et al., 2004). Nevertheless, 30–50% of candidates for CRT have no significant improvement after implantation (Vernooy et al., 2014), which points to the importance of clarifying the criteria for patient selection and optimizing the implantation procedure itself.

Lack of response to CRT is a multifactorial problem associated with variability in individual characteristics, disease patterns, and treatment (Mullens et al., 2009). Combined assessment of multiple factors and individual patient characteristics can improve prediction of response to CRT. With increasing availability of electronic databases, Machine Learning (ML) provides an opportunity to perform such assessment, improving patient selection for therapy (Krittawong et al., 2017; Lopez-Jimenez et al., 2020). Recent studies using ML techniques have achieved impressive results in preoperative clinical data analysis for selecting patients for CRT. Predictive models have been developed to estimate mortality or hospitalization risks from the baseline clinical parameters (Kalscheur et al., 2018; Tokodi et al., 2020, 2021), to assess improvements in EF based on baseline indices and analysis of medical records (Hu et al., 2019) and to stratify patients by an unsupervised learning approach implementing ECG traces (Cikes et al., 2019) and electrocardiography (Feeny et al., 2020). In a recent study (Feeny et al., 2019), Feeny and co-authors using supervised ML approaches selected 9 clinical features (QRS morphology, QRS duration, New York Heart Association CHF classification, LV EF and end-diastolic diameter (EDD), sex, ischemic cardiomyopathy, atrial fibrillation, and epicardial LV lead) that were sufficient to predict patient improvement with fairly high accuracy.

In addition to advances in ML approaches, significant progress has been made in computer modeling of the heart (Auricchio and Prinzen, 2017; Lee et al., 2018). Recent work has shown that patient-specific computer models based on 12-channel ECG and cardiac anatomy measurements are able to reproduce ventricular activation (Potse et al., 2014; Lee et al., 2019; Lopez-Perez et al., 2019; Camps et al., 2021). Moreover, such models may be used to simulate the effect of CRT and study dyssynchrony characteristics (Villongco et al., 2016; Lee et al., 2019).

In recent modeling studies, a combination of cardiac imaging data, personalized models and ML techniques has demonstrated greater accuracy in predicting the propensity for life threatening cardiac arrhythmia in patients with coronary artery disease and cardiac sarcoidosis as compared with current guidelines for ICD implantation (Sung et al., 2020; Shade et al., 2021). Similar hybrid approaches have been used to predict arrhythmia recurrence after pulmonary vein ablation and to target successive ablation procedures in patients with atrial fibrillation (Shade et al., 2020). Our study is another facet demonstrating the relevance of using personalized models as a tool for patient stratification and clinical decision-making.

In this retrospective proof-of-concept study we have developed a new technique to predict CRT response prior to the procedure. First, we developed personalized electrophysiological models simulating ventricular activation and body-surface ECG at the intrinsic activation pattern under left bundle branch block (LBBB) and at BiV pacing mimicking the results of CRT implantation. Then, we used preoperative clinical data along with features derived from clinical image analysis and personalized model simulations to create a supervised multivariable classifier predicting the probability of patient improvement. To the best of our knowledge this has been done for the first time.

The main hypothesis of our study is that personalized model simulations are able to improve ML classification accuracy as compared with pre-operative clinical data alone. Indeed, if the coronary sinus anatomy is available for a patient (which is possible to derive from CT data), one can predict an accessible area for pacing electrode installation and use this area in a personalized ventricular model to simulate BiV pacing. Thus, simulations performed prior to clinical intervention can be used to directly assess the effects of BiV pacing on ventricular activation time, ECG biomarkers and electrical dyssynchrony indices (subject, of course, to the accuracy of model approximation) and hence to estimate whether the goal of the CRT procedure—synchronization of ventricular activation in a particular patient—can be achieved. Importantly, such data can not be derived from pre-operative clinical data itself. Simulated BiV features can be used for patient classification (estimation of CRT response probability) along with other available clinical data and simulated LBBB features.

The study focuses on the following research aims: to assess the contribution of simulated indices derived from personalized modeling to the accuracy of ML predictive models; and to define important clinical and model-derived features in the hybrid dataset for CRT response prediction.

For the first time, we used a hybrid combination of clinical data and model-derived data on ventricular geometry and electrical activation at both intrinsic LBBB pattern and BiV pacing for the development of ML classifier of CRT response. Here, simulated ECG and electrical dyssynchrony features at BiV pacing were selected among the most significant features by means of feature importance algorithms used for classifier development. The ML classifiers on the hybrid dataset outperformed classifiers built upon the clinical data.

2. METHODS

The schematic outline of the research pipeline, including patient cohort selection, clinical indices' acquisition, electrophysiological modeling, feature selection, and machine learning model training, is illustrated in **Figure 1**.

2.1. Clinical Data

2.1.1. Study Population

In this retrospective non-randomized single-center observational study, we enrolled 57 CHF patients on optimal drug treatment who underwent CRT device implantation at Almazov National Medical Research Centre from August 2016 to August 2019. Participants signed approved informed consent. The study protocol was approved by the Institutional Ethical Committee.

The criteria for inclusion into the study were:

1. age over 18;
2. functional class (FC) II-IV of CHF according to the classification of the New York Heart Association (NYHA) at the outpatient stage of treatment;
3. LV EF \leq 35% (Simpson);
4. QRS duration (QRSd) more than 120 ms;
5. sinus rhythm, left bundle branch block (LBBB);
6. optimal drug therapy.

The exclusion criteria were:

1. acute myocardial infarction, transient ischemic attack, acute cerebrovascular accident less than 3 months before the start of the study;
2. patients who were scheduled to undergo myocardial revascularization or heart transplantation during the observation period;
3. congenital and acquired defects, as well as heart tumors, LV aneurysm, if scheduled for surgery during the observation period;
4. active inflammatory and autoimmune diseases of the myocardium;
5. thyrotoxicosis at the time of inclusion in the study;
6. anemic syndrome: blood hemoglobin level less than 90 g/l;
7. diseases limiting life expectancy to less than 1 year.

2.1.2. Data Collection

Patients were evaluated before CRT device implantation and during the follow-up period of 12 months after implantation. Patients underwent investigation according to standard pro-

forma with some additional research methods appropriate for this study.

Standard research methods include:

- clinical examination (complaints, medical history, and physical examination)-before CRT and 1 year after CRT;
- general blood test, biochemical blood test (glucose, potassium, sodium, creatinine, urea, total bilirubin and its fractions, total cholesterol, total protein, AST, and ALT), general urinalysis-before CRT;
- 12-lead ECG-before and 1 year after CRT; ECG monitoring during CRT device programming and during the entire observation period;
- echocardiographic studies before and 1 year after CRT to assess LV reverse remodeling;
- stress tests to exclude/confirm coronary artery disease: stress echocardiography, bicycle ergometry or treadmill test, where clinically indicated;
- coronary angiography, where clinically indicated.

Additional research methods include:

- ECG recording in intrinsic rhythm and under BiV pacing, while programming the CRT device within 7 days after implantation.
- Electrocardiographic imaging using an Amycard system (Amycard, EP Solutions SA, Yverdon, Switzerland). Prior to ECG imaging, a maximum of 224 unipolar body surface mapping electrodes were placed on the patient's torso, followed by computed tomography (CT) imaging of the heart and thorax (Somatom Definition 128, Siemens Healthcare, Germany). Subsequently, the electrodes were connected to the model 01C multichannel electrophysiology laboratory system (Amycard) for continuous ECG recordings during the pacing protocol. CT data were imported into Wave program version 2.14 (Amycard software) to reconstruct 3-dimensional geometry of the torso and heart. Finally, epi/endo ventricle models were manually built with marked active poles of RV and LV leads for bi-ventricular pacing simulations.
- MRI (MAGNETOM Trio A Tim 3 T, Siemens AG or INGENIA 1.5 T, Philips) with contrast (Gadovist or Magnevist) before CRT to detect structural damage of the myocardium.
- Tissue Doppler echocardiography to record ventricle mechanical dyssynchrony. Analysis of interventricular dyssynchrony (IVD) and intraventricular dyssynchrony in the LV (LVD) was performed using biomarkers suggested by Yu et al. (2009). IVD was assessed by the time difference between the start of systolic flows into the aorta and the pulmonary trunk as measured by a pulse-wave Doppler, a value of less than 40 ms was taken as an IVD normal value. LVD was assessed using two biomarkers: dyssynchrony index defined as the temporal difference between the maximal and minimal peak systolic velocities between 12 LV segments ($T_{smax}-T_{smin}$, 105 ms was taken as threshold normal value), and standard deviation in the peak systolic velocities for 12 LV segments ($SD-12$, 34 ms was taken as cutoff value). To determine the peak systolic velocities, the technique of color tissue Doppler ultrasonography was used.

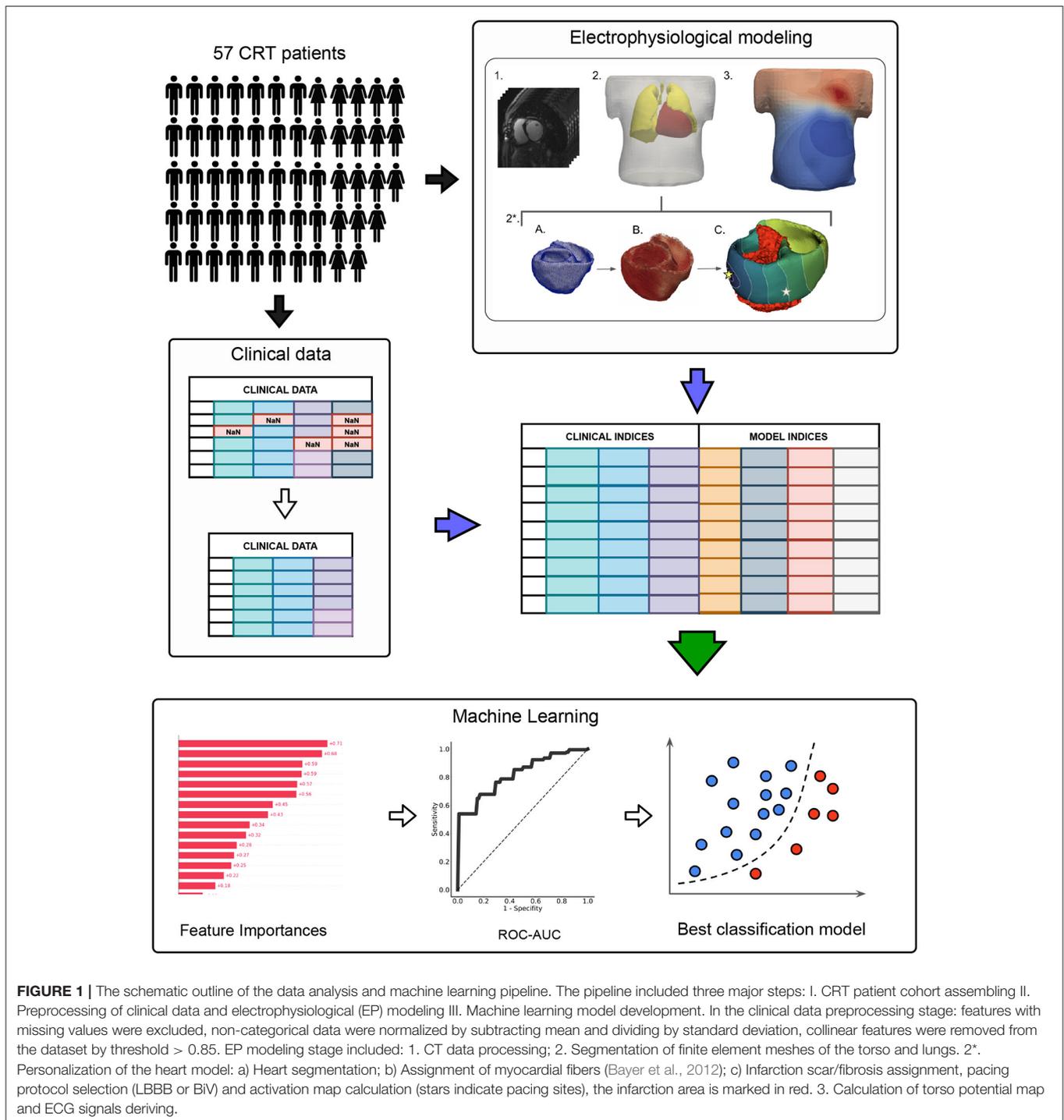


FIGURE 1 | The schematic outline of the data analysis and machine learning pipeline. The pipeline included three major steps: I. CRT patient cohort assembling II. Preprocessing of clinical data and electrophysiological (EP) modeling III. Machine learning model development. In the clinical data preprocessing stage: features with missing values were excluded, non-categorical data were normalized by subtracting mean and dividing by standard deviation, collinear features were removed from the dataset by threshold > 0.85 . EP modeling stage included: 1. CT data processing; 2. Segmentation of finite element meshes of the torso and lungs. 2*. Personalization of the heart model: a) Heart segmentation; b) Assignment of myocardial fibers (Bayer et al., 2012); c) Infarction scar/fibrosis assignment, pacing protocol selection (LBBB or BiV) and activation map calculation (stars indicate pacing sites), the infarction area is marked in red. 3. Calculation of torso potential map and ECG signals deriving.

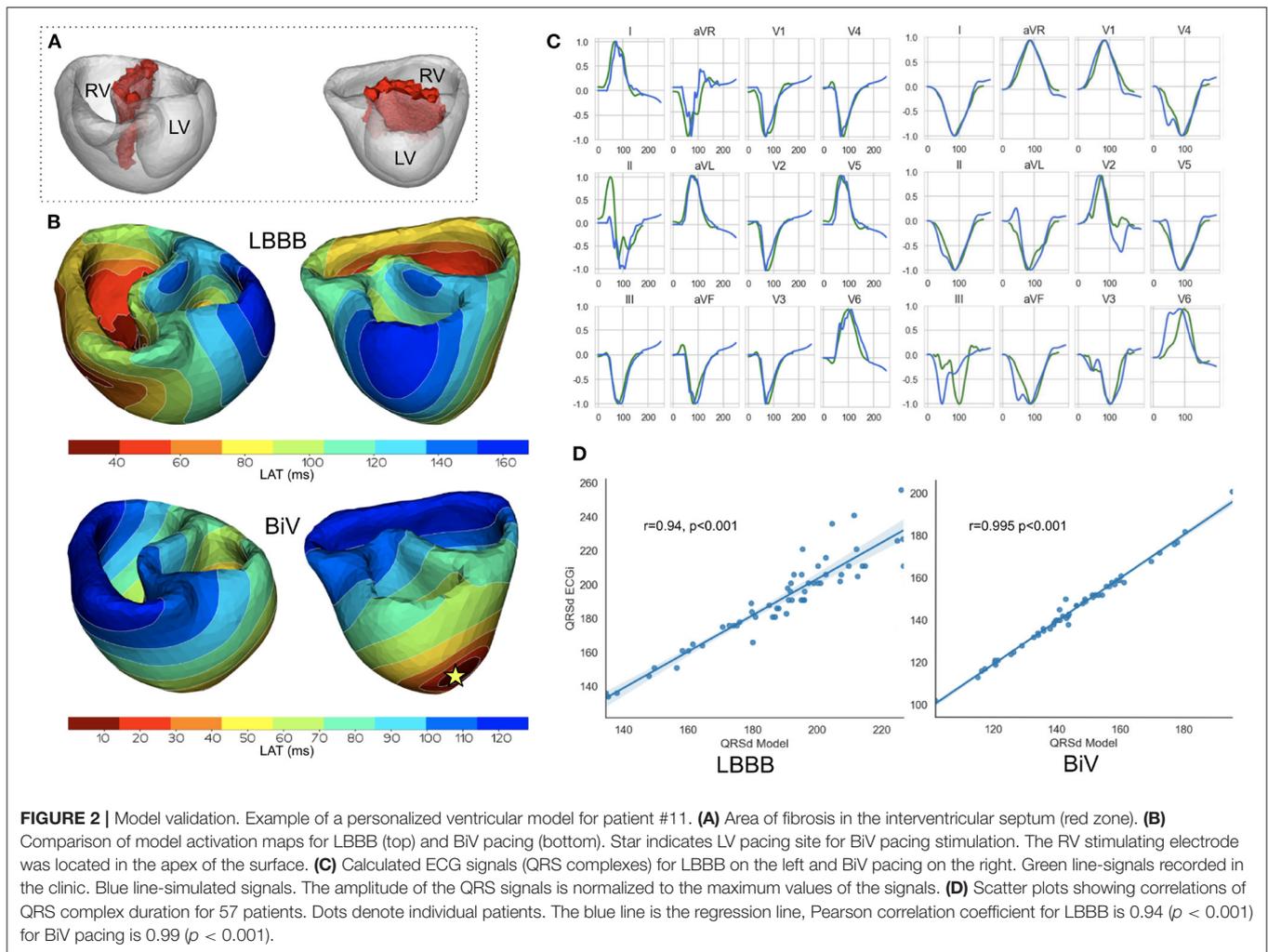
Baseline clinical data for the patients' cohort is presented in **Supplementary Table S1**.

2.2. Personalized Ventricular Models of Electrophysiology

Patient-specific models were generated for each of the 57 cases. Semi-automatic CT segmentation approach helped to extract

torso, lungs and ventricles (**Figure 1** EP modeling, items 1–2). Finite-element meshes were smoothed, refined and merged. Average edge length was 15 mm for torso, 10 mm for lungs and 4 mm for heart.

Then, the LV myocardial tissue in the patient ventricular model was further annotated as either normal tissue or disease-induced remodeling area according to the expert's MRI



examination report from the patient's medical history. The annotation was made as a schematic map of myocardial damage (fibrosis/scar remodeling) using the conventional 17-segment American Heart Association (AHA) model of the LV (Cerqueira et al., 2002), split into three layers, endocardial, mid-myocardial and epicardial, in which damaged regions were highlighted. Each personalized LV geometry computational model was also segmented into 17 x 3 regions (17 segments and 3 layers) according to the AHA scheme. The described areas of scar/fibrosis were labeled on a 17-segment LV AHA model. The scar regions were then simulated as an inexcitable area, and fibrosis regions were associated with a low myocardial conductivity parameter. **Supplementary Figure S1** an example of personalized ventricular model with assigned fibrosis/scar. We also demonstrate a diagram with scar/fibrosis distribution between the segments and the relative volume of the infarct/fibrosis in every segment of AHA LV model (**Supplementary Figure S4**).

The Infarct/fibrosis volume was calculated according to the computational model. The relative volume vs. the myocardial volume was also determined. The compact scar regions were then

simulated as an non-excitable area, therefore such zones were excluded from model calculations. Fibrosis regions were assigned a low constant conductivity value (1% of normal conductivity). **Figure 2A** shows a ventricular geometry model for patient #11 with a zone of intramural fibrosis (red) located in the septum (AHA segments 2,3,8,9).

For every patient-specific ventricular model, the electrical activity in the myocardium and ECG on the body-surface were simulated in several steps. We used an Eikonal model (Keener, 1991) to calculate the activation time at each point of the ventricular mesh. An Eikonal model is widely used to simulate the cardiac activation map as one of the fastest methods (Franzone and Guerri, 1993; Pullan et al., 2006; Pezzuto et al., 2017; Camps et al., 2021). To assign fiber direction at every point of the myocardium, a rule-based approach was used (Bayer et al., 2012). Cardiac tissue was simulated as an anisotropic medium with a conductivity ratio of 4:1 along vs. across the myocardial fibers providing the conduction velocity ratio 2:1 in the fiber and in the transverse direction, respectively. A global value of the conductivity along the fibers was set for the entire myocardial tissue

and fitted against the clinical data as described in the next section.

The QRS complexes in 12-lead ECG on the body surface were computed from the activation time maps simulated by an Eikonal model using the approach for fast phenomenological cardiac models proposed by Pezzuto et al. (2017). In this approach, a predefined cellular action potential is assigned to each model element at corresponding activation time. The potential from the myocardium is transformed on the body surface by utilizing a pseudo-bidomain approach (Bishop and Plank, 2011). To generate action potentials in cardiomyocytes, we employed a widely used cellular ionic model TNNP (ten Tusscher, 2006) for human ventricular cardiomyocytes. We defined the locations of 12-lead ECG electrodes on the torso surface and used simulated QRS complexes of ECG signals for the analysis (Figure 1 EP modeling, items 2-3). Using this approach for ECG calculation, we achieved a high correlation between simulated and experimental QRS durations for our personalized ventricular models (see Figure 2).

2.2.1. Simulation of LBBB Activation Pattern and BiV Pacing

We calculated model-driven indices with reference to two types of ventricular pacing: LBBB activation pattern and BiV pacing.

For LBBB activation, RV sub-endocardial surface was annotated and a Purkinje network was generated using standard parameters from a Costabal model (Sahli Costabal et al., 2016). The right bundle branch was set at about 40 mm length, originating on the intraventricular septum, reaching the RV apex, and then splitting into the Purkinje fiber system of RV (Dobrzynski et al., 2013). His system was isolated from the working myocardium and connected to it only at the ends of the Purkinje fibers through Purkinje-myocardial junction points (PMJs). We set the stimulation time in each PMJ according to the distance to the origin node divided by the conduction velocity in the His-Purkinje system, which we assumed to be 3 mm/ms (Ono et al., 2009).

We used a simulated LBBB ventricular activation map to define an area on late activation time (LAT) in every patient model. LAT zone is frequently considered as a target area for LV electrode installation for the best ventricular synchronization at BiV pacing (Stephansen et al., 2018; Zubarev et al., 2019; Lahiri et al., 2020). We used the distance from LV active pole to LAT zone as one of the model-derived features for training CRT response classifiers.

The location of BiV pacing sites were derived from CT images. Active poles of RV and LV leads were annotated manually in Wave program version 2.14 (Amyscard software). For BiV pacing simulations, we set a zero time delay between the RV and LV pacing sites as programmed in patients.

The activation time at the stimulation points for LBBB and BiV mode of activation was considered as a boundary condition for solving the Eikonal equation.

2.2.2. Personalization of the Electrophysiological Models

Patient-specific ventricular models in both LBBB and BiV pacing protocols were fitted to reproduce individual data from recorded

ECG with the ventricular pacing mode switched off (intrinsic rhythm with LBBB) and switched on (BiV pacing).

For each patient-specific model, we assumed a uniform conductivity in the myocardial tissue in the entire ventricles and solved an optimization problem to find a global conductivity parameter minimizing the discrepancy between simulated and clinical data recorded in the patient in either LBBB or BiV stimulation protocol independently. The post-infarction scar regions were excluded from the model tissue, and a low conductivity of 1% of normal value was assigned to the fibrotic tissue regions when solving the optimization problem. The global conductivity parameter was fitted to minimize the difference between the means of simulated and clinically measured QRSd from the 12-lead ECG recorded in the patient. We used L-BFGS-B algorithm built into SciPy.minimize routine to handle optimization in the model.

Figures 2B,C shows the personalization results for patient #11 with intramural fibrosis located in the septum (AHA segments 2, 3, 8, 9 depicted in red in Figure 2A). Although the model parameters were fitted to minimize the difference between the means of simulated and clinical QRSd, the morphology of the simulated QRS complexes corresponded well with clinical ones (see Figure 2C, blue lines show model signals, green lines show recorded clinical signals). The scatter plots in Figure 2D demonstrate high correlations between simulated and clinical QRSd for both LBBB and BiV modes. The higher correlation coefficient for BiV pacing is explained by precise positioning of the pacing sites derived from CT imaging data, while for LBBB we used a synthetic model of ventricular activation that does not reflect the morpho-anatomical characteristics of the RV conduction system in a particular patient.

2.2.3. Model-Derived Biomarkers of Myocardial Damage, Pacing Site Location, and Myocardial Electrical Activity

Our patient-specific models allowed us to identify several clinically important features affecting ventricular activation. The first group of model-derived indices are based on CT and MRI data coupled with electrophysiology model simulations. Using a digital ventricular model, we were able to define the volume of post-infarction scar and non-ischemic fibrosis and their size relative to the myocardial tissue volume. Knowing RV and LV active poles positions, we measured distance between them (DLvRv). Furthermore, distances from LV pacing site to the infarct/fibrosis area (DLvInfarct) and to the area of LAT (DLvLATZ) under intrinsic rhythm were calculated. When calculating distance biomarkers (see DLvLATZ, DLvLesion and DLvRv in Supplementary Table S1), we solved an isotropic Eikonal equation as a simple method to define the distance from a certain point on the ventricular surface to the border of a specific area. The latter distances mimic distances that can be directly measured from CT or MRI data using a ruler.

The second group of model-derived indices was calculated in LBBB and BiV mode of myocardial activation. We simulated the time activation map for both chambers and 12-lead ECG and calculated the following biomarkers derived from the time-dependent signals for further analysis of CRT response: total ventricular activation time (TAT), maximum QRS complex

TABLE 1 | Model biomarkers.

Index	Definition	Explanation
TAT, ms	$TAT = AT_{max} - AT_{min}$	Total ventricular activation time, AT_{min} (ms) - activation start time, AT_{max} (ms) - late activation time.
QRSd, ms	$QRSd = \max(T_S - T_Q)$	Maximum QRS complex duration in all leads - difference in the Q-S peaks time on the ECG signal.
AT_{RVLV} , ms	$AT_{RVLV} = AT_{maxLV} - AT_{maxRV}$	Interventricular dyssynchrony index - the difference between the time of late activation of LV and RV.
mAT_{STLV}	$mAT_{STLV} = \frac{LVlat_{mean} - ST_{mean}}{TAT}$	LV activation dyssynchrony index, where $LVlat_{mean}$ (ms) - average activation time of LV free wall, ST_{mean} - average activation time of septum.
$IntAV_{STLV}$, ms	$IntAV_{STLV} = \int \frac{AV_{LAT}(t)}{V_{LAT}} - \frac{AV_{ST}(t)}{V_{ST}} dt$	Integral index of LV activation dyssynchrony, $AV_{ST}(t)$ - fraction of myocardial volume activated in septum at time t, $AV_{LAT}(t)$ - fraction of myocardial volume activated in free wall at time t, V_{LAT} - free wall volume, V_{ST} - septum volume.

duration, difference between the total LV and RV activation times (AT_{LVRV}), relative difference between the mean activation times of LV free wall and septum (mAT_{STLV}), integral index of LV free wall and septum myocardial volume activation ($IntAV_{STLV}$). The last three indices characterizing inter- and intraventricular electrical dyssynchrony of myocardial activation were used in work of Villongco et al. (2016).

These simulated features were used as predictions of the effects of BiV pacing on ventricular electrical synchronization. **Table 1** presents a complete list of the simulated characteristics with related definitions and formulas. The average values of all model-derived indices for our enrolled cohort are shown in **Supplementary Table S1**.

Below, changes in feature values under BiV pacing against the LBBB baseline are expressed in relative units. For clinical characteristics, $\Delta X_{CRT} = (X_{CRT} - X_{LBBB}) / X_{LBBB}$, where X_{LBBB} and X_{CRT} are feature values before and after CRT device implantation, respectively. For simulated indices, $\Delta X_{BiV} = (X_{BiV} - X_{LBBB}) / X_{LBBB}$, where X_{LBBB} and X_{BiV} are feature values in the LBBB activation mode and BiV pacing. For the indices that are initially showed in relative units, e.g., EF and mAT_{STLV} , the response to pacing is expressed as an absolute increment in the LBBB value: $\Delta EF_{CRT} = EF_{CRT} - EF_{LBBB}$, and $\Delta mAT_{STLV BiV} = mAT_{STLV BiV} - mAT_{STLV LBBB}$.

For some features, we also used normalized characteristics, i.e., the ratio of the value to the myocardial tissue volume (MTV). Note, MTV should not be confused with the volume of the cavity inside the ventricle. We can evaluate MTV using a digital model of the ventricular geometry based on CT images. When using such normalization, the volume of the preserved myocardium only is taken into account, without allowing for the infarct area. Since the scar tissue is not excited and does not contract, it is excluded from the volume of the active ventricular myocardium. Normalized features give a characteristic's values per unit volume of the myocardium (analogous to the values per mass unit of the myocardium). For instance, TAT/MTV indirectly reflects a reciprocal value of the average velocity of myocardium activation in the ventricles.

2.3. A Predictive Model of Response to CRT Based on Preoperative Clinical Data and Electrophysiology Model Simulations

For classifier development, we applied several supervised machine learning (ML) approaches to identify an optimal set of features and learning algorithm combination showing the

best performance characteristics on hybrid data for our patient cohort. The hybrid dataset for building the classifier contained clinical and model-derived features as described above. At the preprocessing step, features with missing values were excluded. Non-categorical data were normalized by subtracting the mean and dividing by standard deviation. Collinear features were also removed from the dataset by threshold > 0.85 .

Several criteria for CRT response definition were used for classification. The primary criterion for responders was more than 10% increase in LV EF (EF10) (Feeny et al., 2019). The following criteria were also considered: a reduction in ESV $> 15\%$ (ESV15, see Foley et al., 2009; Park et al., 2012); a 5 and 15% increase in LV EF (Feeny et al., 2019), and combined EF10 and ESV15.

We evaluated several classification algorithms: logistic regression (LR), linear discriminant analysis (LDA), support vector machine (SVM) with linear kernel, random forest (RF) classifier; each evaluated in combination with three different feature sets obtained by feature selection methods. The following algorithms were used for feature selection: random forest mean decrease accuracy (MDA), univariate statistical testing (UST, two-sample *t*-test for continuous variables and chi-squared test for categorical variables), and L1-based feature selection (L1, based on weights of LR). Features were selected in a cross-validation loop for each subset. The top 8 features chosen by the algorithms were used to construct the classifiers.

Feature selection and training of classification algorithms was done using a Leave-One-Out cross-validation loop. Within the loop, the ML classifier score for each test fold (each consisting of just one observation) was calculated. These ML scores were combined into one set to build the receiver operating characteristic (ROC) curve and to calculate the area under the ROC curve (AUC). The highest-performing combination of the classifier with feature selecting algorithm was chosen to develop the final classifier.

In addition, we used repeated stratified five-fold cross-validation with 1,000 iteration in order to be confident in assessing the quality of the classifier. We quantified the classification performance of each feature set-algorithm combination with ROC AUC across all folds and iterations. The classifier with the highest ROC AUC was selected as the final classifier for our hybrid dataset.

2.3.1. Software

Cardiac electrophysiology was modeled with the help of software written at the Institute of Immunology and Physiology UB RAS

based on FENICS library (for solving PDE problems) (Logg and Wells, 2010) and VTK (for working with meshes). For the machine learning pipeline (see **Figure 1**): classifier development, statistical modeling, feature selection, cross validation, and ROC-AUC calculation we used the sklearn library.

2.3.2. Statistics

Detailed analysis was performed using the IBM SPSS Statistics 23.0.0.0 software package (USA). For qualitative data, the frequency and percentage of total patients in the cohort were calculated. Quantitative data are presented as mean \pm standard deviation. Comparisons between two dependent groups were made using Wilcoxon's test for quantitative data and McNemar's test for qualitative data. Nonparametric Friedman's two-way ANOVA was applied to compare related groups. Comparison between two independent groups was carried out using the Mann-Whitney test for quantitative data and Pearson's chi-square test for qualitative data. Feature dependence was assessed using the Spearman rank correlation test. The critical level of statistical significance was taken equal to 0.05.

3. RESULTS

3.1. Responders vs. Nonresponders: Analysis of Clinical Data Before and After CRT Device Implantation and Model Simulations in LBBB and BiV Pacing

We found an average positive response to BiV pacing in all clinical indicators and corresponding simulated indexes of the CRT outcome in the entire patient cohort (a summary of statistics for clinical data, CT/MRI derived data and model-driven biomarkers is presented in Section S.1, **Supplementary Table S1**). High variability of the effects of BiV pacing on biomarkers in both the clinical and simulated data suggest a significantly nonuniform output among the patients. Therefore, the patients were classified into two groups of responders and nonresponders to the therapy.

We have used several conventional criteria to classify responders and nonresponders to CRT in the patient cohort based on clinical data on the post-operative LV reversed remodeling. Primary classification was defined by a higher than 10% increase in the LV EF for responders ($\Delta EF_{CRT} > 10\%$ referred hereafter as EF10 criterion). This criterion was used in clinical studies, and allowed us to compare qualitatively the results of our predictive models for CRT response with the findings reported recently by Feeny et al. (2019). Surprisingly, the 10% cutoff for EF improvement in responders is close to the average EF increase of $9 \pm 8\%$ observed in our patient cohort. Classification results based on other CRT response definitions are described in the **Supplementary Materials** and discussed in the section 4.

Table 2 summarizes the clinical and model-derived variables in the groups with or without LV EF improvement according to the EF10 criterion. In our patient cohort, 23 (40%) patients demonstrated an improved EF (referred to as CRT responders) and 34 (60%) patients were classified as nonresponders. The ratio

seems biased toward nonresponders, but we have intentionally raised the LV EF improvement threshold in order to be more confident in predicting true positive responses. Average EF is raised in both groups, and the increase is significantly higher in the responders vs. nonresponders ($17 \pm 5\%$ vs. $3 \pm 5\%$, respectively). The EF improvement after CRT is accompanied by a prominent ESV reduction by $47 \pm 19\%$ in the responder group against an insignificant diminishing by $9 \pm 37\%$ in nonresponders. Similarly, a much higher average EDV reduction is seen in the responders due to LV postoperative reverse remodeling after CRT. Although the average QRSd is decreased, no statistical significance between the groups was found. No difference in the CRT effect on the mechanical dyssynchrony indices was found as well.

In consistency with the clinical data, the model simulations revealed a decrease in both TAT and QRSd under BiV pacing in each of the two sub-populations of models (**Figure 3**). The electrical dyssynchrony indices also reveal a prominent decrease in the groups, with the highest reduction in the inter-ventricular dyssynchrony index AT_{RVLV} (**Figure 3**). Meanwhile, no difference in the relative decrease in the indexes between the responder and nonresponder groups was observed (**Table 2**).

Analyzing CT/MRI derived geometry indexes, we figured out no difference in the relative volume of infarct/fibrosis in LV myocardial tissue between the groups (**Table 2**). At the same time, we found a shorter distance from the LV pacing site to the damaged zone in the nonresponder group (28 ± 27 mm in nonresponders vs. 45 ± 28 mm in responders), suggesting less effective pacing of the normal tissue in nonresponders. Interlead spacing does not statistically differ between groups. No difference in the distance from LV pacing site to the LAT area in LBBB activation mode was found as well.

It is of note that most of the individual biomarkers in the intrinsic LBBB activation pattern derived from either clinical, or CT/MRI, or simulated data do not show a significant difference in the distribution between the responder and nonresponder groups. This means that no one single index could be considered as a diagnostic feature for preoperative classification (**Table 2**).

Among the pre-operative clinical data, two features, i.e., LV EF_{LBBB} and the inter-ventricular mechanical dyssynchrony index IVD_{LBBB} , displayed differences between the groups classified according to the EF10 criterion. Here, LV EF demonstrated a bit higher average value along with a bit lower value of IVD in the nonresponders than in responders. This is consistent with a low negative correlation between EF_{LBBB} before and ΔEF_{CRT} after implantation ($r = -0.48$, $p = 0.031$, see **Supplementary Figure S7**). However, high EF variation in each group comparable with the difference between the group averages did not allow us to find a valid threshold separating the groups. The average accuracy of the Logistic Regression classification with One-Leave-Out cross-validation based on EF_{LBBB} was only 0.62 with rather low values of both sensitivity at 0.69 and specificity at 0.55. A low positive correlation was also found between IVD_{LBBB} and ΔEF_{CRT} ($r = 0.32$, $p = 0.029$), suggesting its possible predictive power for CRT response. However, we did not have IDV and other mechanical dyssynchrony indices for all

TABLE 2 | Clinical, imaging, model data and predictive model scores for responders and nonresponders defined by EF10 criterion.

Variable	Patient cohort <i>n</i> = 57					
	Responders <i>n</i> = 23 (40%)			Nonresponders <i>n</i> = 34 (60%)		
Clinical data						
Gender (male/female)	15/8			23/11		
Age, year	64 ± 6			63 ± 7		
BMI	27 ± 5			30 ± 5#		
IHD/DCM	14 (61%)/9(49%)			22 (65%)/12(35%)		
History of AF	4 (17%)			8 (24%)		
FC CHF :	LBBB	CRT	Δ, %	LBBB	CRT	Δ, %
I	0 (0%)	7 (30%)*	7	0 (0%)	3 (9%)	3
II	12 (52%)	12 (52%)	0	12 (35%)	19 (56%)**	7
III	11(48%)	2 (8%)*	-9	22 (65%)	4 (12%)**	-18
QRSd, ms	192 ± 20	143 ± 14**	-25 ± 11	190 ± 26	145 ± 21**	-22 ± 16
Echocardiography data						
EDV, ml	301 ± 69	196 ± 68**	-33 ± 22	290 ± 106	263 ± 138##	-7 ± 33##
ESV, ml	231 ± 59	119 ± 48**	-47 ± 19	207 ± 87	185 ± 118*##	-9 ± 37##
EDD, mm	74 ± 8	62 ± 10**	-16 ± 10	73 ± 7	69 ± 9 ***#	-5 ± 8 ##
ESD, mm	64 ± 9	48 ± 13**	-26 ± 17	62 ± 9	57 ± 10 ***#	-7 ± 13##
EF, %	23 ± 5	40 ± 6**	17 ± 5	29 ± 6##	32 ± 7***#	3 ± 5##
IVD, ms (<i>n</i> = 34)	76 ± 17	46 ± 22**	-38 ± 29	63 ± 19##	33 ± 14**	-44 ± 27
ΔTs, ms (<i>n</i> = 34)	82 ± 35	76 ± 34*	-20 ± 39	87 ± 44	58 ± 33*	-12 ± 78
SD12, ms (<i>n</i> = 34)	31 ± 14	27 ± 12	-20 ± 40	33 ± 16	23 ± 12**	-15 ± 74
CT/MRI data						
MTV,ml	332 ± 142			377 ± 143		
InfarctV, ml	45 ± 39			54 ± 39		
InfarctV/MTV	0.14 ± 0.08			0.16 ± 0.13		
DLvRv,mm	108 ± 23			105 ± 25		
DLvLATz,mm	44 ± 16			58 ± 27		
DLvInfarct,mm	45 ± 28			28 ± 27#		
Model data						
	LBBB	BiV	Δ, %	LBBB	BiV	Δ,%
TAT, ms	269 ± 109	141 ± 31**	-45 ± 18	246 ± 130	138 ± 46**	-45 ± 24
QRSd, ms	192 ± 21	143 ± 14**	-30 ± 12	187 ± 24	152 ± 28**	-22 ± 20
AT _{RVLV} , ms	103 ± 65	26 ± 26**	-75 ± 21	95 ± 79	20 ± 21**	-76 ± 27
IntAV _{STLV} , ms	101 ± 57	34 ± 15**	-51 ± 36	106 ± 59	33 ± 16**	-53 ± 52
mAT _{STLV}	0.36 ± 0.10	0.29 ± 0.14	-6 ± 20	0.36 ± 0.09	0.27 ± 0.13**	-9 ± 17
Predictive model scores						
Score by Feeny et al. (2019)	0.63 ± 0.20			0.55 ± 0.23		
MLCD score (EF10)	0.47 ± 0.23			0.37 ± 0.24		
MLHD score (EF10)	0.58 ± 0.25			0.29 ± 0.22##		

Mean ± SD.

p* < 0.05, *p* < 0.01 LBBB vs. CRT or LBBB vs. BiV. Comparisons between two dependent groups were made using Wilcoxon's test for quantitative data and McNemar's test for qualitative data.

#*p* < 0.05, ##*p* < 0.01 Responders vs Nonresponders. Comparison between two independent groups was carried out using the Mann-Whitney test for quantitative data and Pearson's chi-square test for qualitative data.

Δ - Average change in indicator $\Delta X = X_{CRT} - X_{LBBB} / X_{LBBB}$ or $\Delta X = X_{BiV} - X_{LBBB} / X_{LBBB}$. Δ is calculated as the absolute difference for normalized values (EF and mAT_{STLV}) and FC.

BMI, Body mass index; IHD, Ischemic heart disease; DCM, Dilated cardiomyopathy; AF, Atrial Fibrillation; FC CHF, functional class of congestive heart failure; IVD, interventricular dyssynchrony; ΔTs, maximum temporary difference in peak systolic velocities between 12 LV segments; SD12, standard deviation of the peak systolic velocities of 12 LV segments; MTV, myocardial tissue volume; LAT, late activation time; TAT, total ventricular activation time; QRSd, maximal duration of QRS complex on 12 leads; AT_{RVLV}, difference of total LV and RV activation time; IntAV_{STLV}, integral index of LV free wall and septum myocardial activation dyssynchrony; mAT_{STLV}, difference between mean activation time of LV free wall and septum; MLCD score (EF10), ML score on the clinical data for EF10 criterion; MLHD score (EF10), ML score on the hybrid data for EF10 criterion.

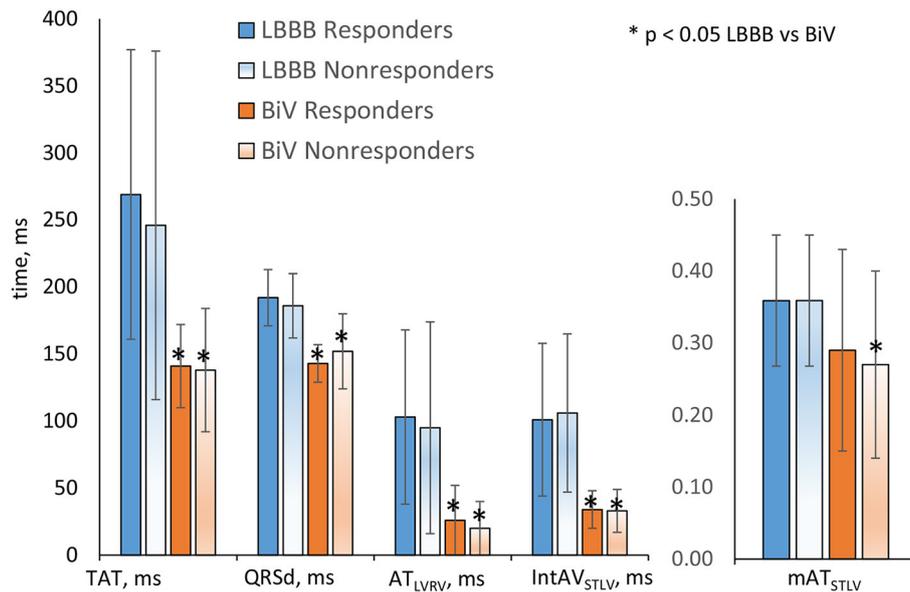


FIGURE 3 | Simulation features in the LBBB activation mode and under BiV pacing for responders and nonresponders. Bar indicates mean. Error bar is SD. Comparisons between two dependent groups (LBBB vs. BiV) were made using Wilcoxon's test. No difference in the indexes between the responder and nonresponder groups was observed. * $p < 0.05$ LBBB vs. BiV.

57 patients in our cohort and, therefore, decided against using them in further analysis (see section 5).

Although in this proof-of-the-concept study some data used for model building were recorded after operation, we consider all CT/MRI and model-derived features as potentially pre-operative because all of them can actually be assessed before operation (see also section 5. on this issue). Among the data derived from CT/MRI, we found only the distance between the LV pacing site and the area of myocardial damage showing a significant difference between the responders and nonresponders (Table 2). However, this index did not show a significant correlation with either ΔEF_{CRT} (see Supplementary Figure S7) as well.

In consistency with the absence of difference between clinical QRSd in responders and nonresponders, none of the simulated electrophysiological biomarkers showed any significant difference between groups both in the LBBB mode of activation and under BiV pacing either (Figure 3 and Table 2), which also did not allow them to be considered as individual classifying features.

Our dataset analysis suggested a hypothesis that the only combination of the clinical and MRI/CT derived biomarkers that can be evaluated before operation together with predictions on the BiV response simulated using a personalized ventricular model may increase the predictive power of such a hybrid dataset for patient classification.

3.2. Predictive Models of CRT Response Built on Hybrid Dataset of Clinical Data Before Operation and Personalized Model Simulations at LBBB and BiV Pacing

We used the hybrid input dataset containing 57 data entries with features derived from clinical data recorded prior to operation,

CT/MRT derived data and simulated features calculated using personalized models of ventricular excitation in LBBB and BiV pacing activation modes for every patient from our cohort as described in the previous sections. The complete list of features fed to the feature selection algorithms when developing CRT response classifiers is shown in Figure 4 (right) in descending order of the feature importance. We trained supervised classifiers using an EF10 criterion ($\Delta EF > 10\%$) of CRT response. To choose the best classifier, we compared 4 different classification models (classifiers) with Leave-One-Out and five-fold cross-validation and 3 different feature selection methods inside a cross-validation loop. A summary of the model ROC AUC used to characterize the quality of the trained models is shown in Table 3. It is seen that average ROC AUC vary from a smallest value of 0.7 to the best one of 0.82 obtained for SVM and LDA classifiers with Univariate approach for feature selection.

Figure 4 (left) shows a ROC curve for the best SVM classifier trained for the EF10 response criterion. Table 4 summarizes the classifier characteristics. The best SVM classifier for CRT response demonstrates a high accuracy of 0.82, sensitivity of 0.85, and specificity of 0.78.

ML scores generated by the best SVM classifier correlate with post-operational improvement in the EF ($r = 0.46$, $p < 0.001$, see Figure 5). Moreover, the distributions of the average scores in the responder and nonresponder groups in our patient cohort significantly differ between each other with a significantly higher average score in the responder vs. nonresponder group (0.58 ± 0.25 vs. 0.29 ± 0.22 , $p < 0.01$, see Figure 6 and Table 2). A corresponding score threshold of 0.46 was defined for the responders in our patient cohort for the best classifier according to the EF10 response definition.

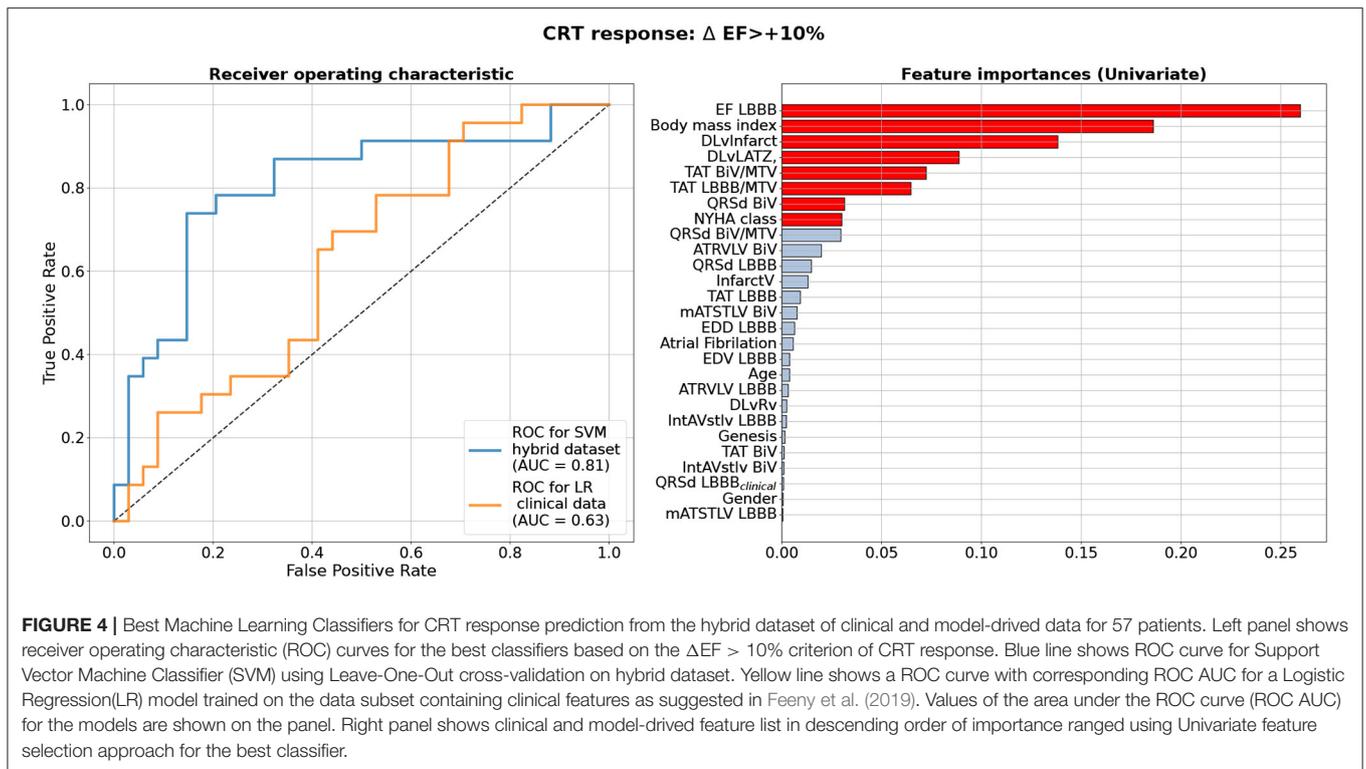


FIGURE 4 | Best Machine Learning Classifiers for CRT response prediction from the hybrid dataset of clinical and model-driven data for 57 patients. Left panel shows receiver operating characteristic (ROC) curves for the best classifiers based on the $\Delta EF > 10\%$ criterion of CRT response. Blue line shows ROC curve for Support Vector Machine Classifier (SVM) using Leave-One-Out cross-validation on hybrid dataset. Yellow line shows a ROC curve with corresponding ROC AUC for a Logistic Regression (LR) model trained on the data subset containing clinical features as suggested in Feeny et al. (2019). Values of the area under the ROC curve (ROC AUC) for the models are shown on the panel. Right panel shows clinical and model-driven feature list in descending order of importance ranged using Univariate feature selection approach for the best classifier.

TABLE 3 | Comparison of ROC AUC for different Machine Learning classifiers using leave-one-out and five-fold cross-validation and different feature selection algorithms for EF10 criterion of CRT response.

Classifier	Leave-one-out			Five-fold		
	Feature selection method			Feature selection method		
	L1	MDA	Univariate	L1	MDA	Univariate
Logistic regression	0.76	0.78	0.80	0.74 ± 0.15	0.76 ± 0.14	0.80 ± 0.14
Linear discriminant analysis	0.73	0.76	0.82	0.71 ± 0.15	0.76 ± 0.15	0.80 ± 0.14
Support vector machine	0.70	0.73	0.82	0.72 ± 0.15	0.75 ± 0.15	0.80 ± 0.14
Random forest	0.73	0.72	0.73	0.72 ± 0.16	0.70 ± 0.16	0.72 ± 0.16

L1, Logistic Regression feature selection; MDA, Mean Decrease Accuracy; Univariate, Univariate statistical testing: two-sample t-test for continuous variables and chi-squared test for categorical variables. Bold text indicates the best classifier.

Figure 4 (right) shows a ranged list of feature importances selected by the SVM classifier trained on the entire dataset for the EF10 CRT response criterion. Eight most important features colored in red were selected for the final classifier. The pre-operative EF_{LBBB} showed the highest importance among other inputs, which is in line with our findings on the correlation between ΔEF_{CRT} and EF_{LBBB} . The other two of the three clinical features contributing to the CRT response were BMI and NYHA stage. Therefore, the majority of the selected features were indices derived from CT/MRI and simulated features in the LBBB and BiV modes of activation. In particular, the distance between the LV pacing

site and the infarct/fibrosis area was the third in the feature importance range, and a combination of TAT/MTV_{LBBB} , TAT/MTV_{BiV} and $QRSd_{BiV}$ showed the highest importance among simulated features. Corresponding coefficients at the input variables in the terms of the best LR classifier are given in **Supplementary Table S2**.

The yellow line in **Figure 4** (left panel) shows the ROC curve for an LR classifier trained on the clinical data only according to the EF10 criterion. The sub-set of clinical features used here for CRT response prediction was the same as selected in the article by Feeny et al. (2019) for their best LR classifier (see the complete feature list and corresponding coefficients at the input variables

TABLE 4 | Performance of the classifiers on hybrid vs. clinical data.

SVM on hybrid dataset				
$\Delta EF > +10\%$, cutoff = 0.46				
	Responder		Non-responder	
Model responder	18		5	
Model Non-responder	5		29	
Accuracy	Sensitivity	Specificity	ppv	npv
0.82	0.85	0.78	0.78	0.85
LR on hybrid dataset				
$\Delta EF > +10\%$, cutoff = 0.46				
	Responder		Non-responder	
Model Responder	15		8	
Model Non-responder	7		27	
Accuracy	Sensitivity	Specificity	ppv	npv
0.74	0.65	0.79	0.68	0.77
LR on clinical dataset				
$\Delta EF > +10\%$, cutoff = 0.51				
	Responder		Non-responder	
Model Responder	8		12	
Model Non-responder	15		22	
Accuracy	Sensitivity	Specificity	ppv	npv
0.53	0.65	0.35	0.40	0.59
Response score by Feeny et al. (2019)				
$\Delta EF > +10\%$, cutoff = 0.62				
	Responder		Non-responder	
Model Responder	13		15	
Model Non-responder	10		19	
Accuracy	Sensitivity	Specificity	ppv	npv
0.56	0.56	0.57	0.46	0.65

SVM, Support Vector Machine; LR, Logistic Regression; ppv, positive predictive value; npv, negative predictive value.

in **Supplementary Table S2**). The average ROC AUC for this predictive model appears to be 0.63 for our patient cohort, with an average accuracy of 0.53, sensitivity of 0.65, and specificity of 0.35 (**Table 4**), which are much lower than the characteristics of the ML model trained on the hybrid input dataset containing a combination of clinical and model-driven features. Note that this AUC is close to the AUC value of 0.62 we obtained for the LR classifier trained on EF_{LBBB} only, suggesting that the rest of the clinical information does not contribute essentially to the model predictions.

We compared also the accuracy of EF10 improvement predictions from our ML classifier on the hybrid data with the accuracy of predictions based on our patients' clinical features fed into a "ML score calculator" presented in Feeny et al. (2019) (**Table 4**). This predictor showed an accuracy of 0.56, sensitivity of 0.56, and specificity of 0.57 on our 57 patient dataset, which are similar with the performance of the LR classifier trained on the clinical data, but much lower than the performance of our ML classifiers on hybrid data.

Figure 6 shows average ML scores generated by the classifier on the hybrid data for the entire patient cohort and for the responder and nonresponder groups according to the EF10

definition in comparison with the ML scores predicted by the LR classifier trained on the clinical data from our patient cohort and those from the calculator by Feeny et al. (2019). The average ML score from Feeny et al. (2019) on the entire patient cohort is seen to be higher than our ML scores, explaining lower rates of true positive and true negative predictions from the calculator on our patient cohort. Moreover, the only classifier on hybrid data generates significantly higher ML scores in the responder vs. nonresponder group, suggesting its higher predictive performance. In contrast, the average ML scores did not differ between responders and nonresponders according for the LR classifier on the clinical data (0.47 ± 0.23 vs. 0.37 ± 0.24 , $p = 0.111$) and for the Feeny's calculator from Feeny et al. (2019) (0.63 ± 0.20 vs. 0.55 ± 0.23 , $p = 0.213$) in our patient cohort (see also **Table 2**).

These results clearly highlight the significance of model-driven features for CRT response prediction.

4. DISCUSSION

The researchers sought ways to predict CRT response for ensuring more effective patient stratification and different

outcome end-points for improving state, increasing survival period and preventing adverse effects (Lahiri et al., 2020). Despite of intensive research performed in the field, the fraction of patients with low response to the therapy remains as high as 30–50% depending on which criteria are used for assessing CRT outcome. New artificial intelligence and ML based approaches to data analysis have been extensively used in attempts to increase the accuracy of patient differentiation (Kalscheur et al., 2018;

Feeny et al., 2019, 2020; Tokodi et al., 2021). Computational models based on clinical data are also employed to identify mechanisms responsible for the poor efficacy and develop approaches improving CRT outcomes (Lumens et al., 2015; Huntjens et al., 2018; Lee et al., 2018; Isotani et al., 2020). Recently, a new trend has emerged in this research area, which uses a combination of clinical and model data together with ML for solving challenging medical problems (Aronis et al., 2021; Heijman et al., 2021). As far as we know, there have been no reports of *in-silico* studies involving a hybrid approach to predict CRT response in a cohort that would combine a dataset of patient-specific features derived from clinical measurements and simulations on personalized ventricular models.

4.1. Improvement of Classification Models Built on Hybrid Data vs. Predictors on Clinical Data

In this study, we combined the MR/CT-imaging and model derived features with pre-operative clinical data used conventionally to characterize patient's state in a hybrid dataset for building predictive models of CRT response in the patients by ML techniques. A sub-set of simulated features containing TAT, QRSd and three electrical dyssynchrony indices generated by every of 57 patient-specific electrophysiology models under LBBB and BIV pacing was used as an input to ML algorithms. The personalized models were also used to define LAT zone in the LV under LBBB mode of activation and to calculate the distances between the pacing sites, and from the LV pacing site to the LAT zone and to the LV infarct/fibrosis area, which were also used as input features for ML classifiers. The basic hypothesis of our study was that model-driven simulations of the response to BiV pacing may essentially enhance the predictive power of the hybrid dataset for CRT response evaluation.

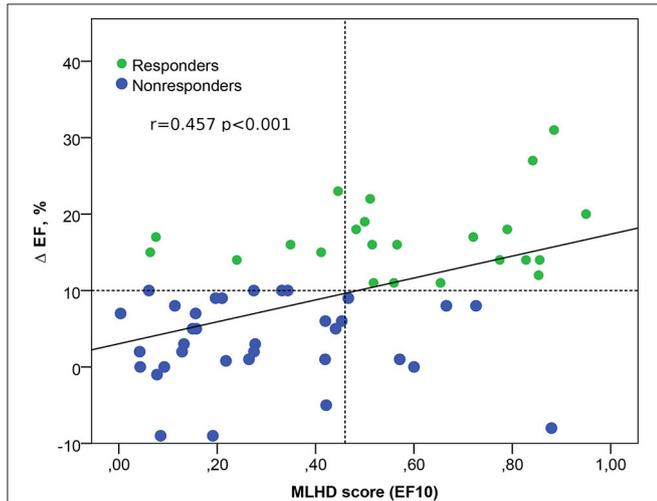


FIGURE 5 | Relation between the ML score on the hybrid data (MLHD score) for EF10 criterion and the post-operational change in the EF. Solid line - regression line $\Delta EF = 3 + 14 \text{ MLHD score}$; horizontal dotted line shows a 10% threshold for LV EF improvement; vertical dotted line is a MLHD score cutoff of 0.46 for responders; r is the Spearman correlation coefficient; p is the significance for the group difference.

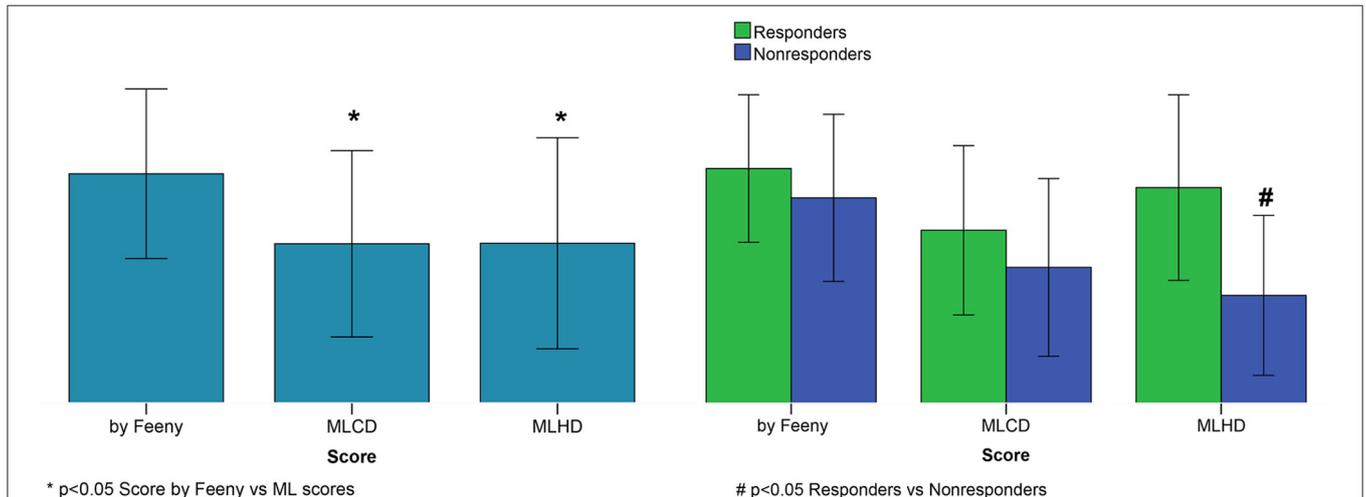


FIGURE 6 | CRT response scores. Left panel: Average scores. Right panel: Average scores for responders and nonresponders. Score by Feeny et al. (2019); MLCD–ML score on the clinical data for EF10 criterion; MLHD–ML score on the hybrid data for EF10 criterion. Bar indicates mean. Error bar is SD. Nonparametric Friedman's two-way ANOVA was used to compare related groups (Score by Feeny vs. ML scores). Comparison between two independent groups (responders vs. nonresponders) was performed using the Mann-Whitney test.

Despite the rather small size of the dataset used for ML classifier development (57 entries in the entire dataset), we were able to obtain ML classification models achieving high accuracy in predicting the response to CRT (see **Figure 4** and **Table 3**). The ROC AUC value for the best SVM classifier is as high as 0.82 for $\Delta EF > 10\%$ cutoff for responders.

The most significant result of our study is that our best classification models built on the hybrid dataset outperformed the ML classifier trained on the pre-operative clinical data only (see **Figure 4** and **Table 4**). In the latter, we used a subset from the same 57 patient dataset containing 9 clinical features for each patient. The features were selected in recent study by Feeny et al. (2019) as most important for training the best LR model based on the clinical data from a thousand of patients. So, we used the same clinical features to train similar LR classifier on the data from our patient cohort. The best model built on the clinical data demonstrated a ROC AUC of 0.63, and the accuracy, sensitivity and specificity much lower than those for the classifiers built on the hybrid dataset (see **Figure 4** and **Table 4**). Then we used the same 9 clinical features for every 57 patients as a testing dataset to feed to the ML score calculator provided by Feeny et al. (see the **Supplementary Materials** in Feeny et al., 2019). It showed an accuracy of 0.56 for our patient cohort. The performance of the ML score calculator tested on the clinical data from our 57 patients is similar with the performance of the LR classifier we trained on the same data. Both classifiers showed lower performance as compared with the classifiers trained on the hybrid data from the same 57 patients. Moreover, our hybrid data classifier outperformed classifiers reported in Feeny et al. (2019) (see **Table 3** *ibidem*), which were trained on different sets of clinical data from about thousand of patients. Note, the clinical data classifiers on a large dataset demonstrated higher metrics than those built on 57 clinical data inputs. This supports our expectation of further improvement of the hybrid data classifier with data-set extension. Therefore, we may conclude that our ML classifiers built on the combination of clinical and model-derived features significantly improve CRT prediction quality with higher accuracy, sensitivity and specificity.

In addition, we compared the average ML CRT response scores in the responder and non-responder groups provided by the best SVM classifier on hybrid data, the LR classifier on the clinical data and that provided by the ML score calculator from Feeny et al. (2019) (see **Figure 6**). Noteworthy, for the SVM classifier based on the hybrid data, we found a significantly higher average score in the responders vs. nonresponders confirming the predictive power of the ML model. In contrast, the average ML scores predicted by the LR classifier on the clinical data and calculator from Feeny et al. (2019) did not differ between responders and nonresponders in our patient cohort (see **Table 2**).

Therefore, our results clearly show significant advantages ensured by the use of hybrid data combining clinical data with simulated features from personalized electrophysiology models for building ML predictive models of CRT response.

4.2. Feature Selection for Classification Models From Hybrid Data

During classifier development, we tested several feature selection methods for different classifiers and different numbers of features to define the final model with best characteristics (see **Figure 4**). Note that we did not predetermine input features for classifiers based on prior analysis. Instead, the features were automatically selected inside the cross-validation loop as described in section 2. The final feature lists selected for the best predictive models contain 8 inputs. Importantly, the most important feature set contained fewer clinical features compared with model-derived ones. In consistency with ESC guidelines on the significance of pre-operative baseline LV EF_{LBB} for CRT response, it was selected as the most important feature for the classification model based on the EF10 definition (see **Figure 4**). Interestingly, BMI was selected at the second position in the feature chart. The latter result is in line with study by Hsu et al. (2012), who demonstrated that BMI < 30 kg/m² predicted LV EF super-response.

We also tested the importance of model-driven characteristics extracted from the CT/MRI data coupled with model simulations. In our study, LV myocardial damage volume (both absolute and relative to the survival myocardium volume) did not reveal high importance by itself, but the distance from the LV pacing site to the infarct/fibrosis area was selected as the third most important feature for classifiers (see **Figure 4**). We found no significant correlations between this distance and the post-operative values of LV EF improvement ΔEF_{CRT} or ESV reduction ΔESV_{CRT} (see **Supplementary Figure S7**). However, the role of the distance from the LV pacing site to the infarct/fibrosis area in CRT response prediction was supported by a positive correlation between the ML score and the distance ($r = 0.445$, $p = 0.001$). As expected, much higher average distance in the responder vs. nonresponder group (45 ± 28 vs. 28 ± 27 mm, $p = 0.02$, see **Table 2**) was found.

Our findings are consistent with the results of clinical studies which assessed the significance of myocardial infarct size for CRT response. The extent of scar core and gray zone was automatically quantified using cardiac MRI analysis (Nguyễn et al., 2018a). The highest percentage of CRT response was observed in patients with low focal scar values and high QRS area before operation. Such area was calculated using vector-cardiography. In study by Marsan et al. (2009) MRI was performed in candidates to derive LV mechanical dyssynchrony and the extent of scar tissue to predict CRT response. Higher LV dyssynchronies were strongly associated with echocardiographic response to CRT, while the total extent of scar correlates with non-response. Importantly, a univariable logistic regression analysis showed that the presence of a match between the LV lead position and a transmural scar was also significantly associated with non-response to CRT. The location of scar in the posterolateral region of the LV, which is empirically thought to be a target site for LV lead implantation, was associated with lower response rates following CRT (Chalil et al., 2007). In study by Pezel et al. (2021), no difference was found in presence and extent of scar between CRT responders and non-responders. However, in non-responders, the LV lead

was more often over an akinetic/dyskinetic area suggesting the presence of tissue lesions, a fibrotic area, or an area with myocardial thickness < 6 mm.

As seen in **Figure 4**, the distance from the LV pacing site to LAT zone was selected as a 4-th feature in the importance list for EF10 definition of CRT response. Accordingly, we revealed a low negative correlation between the distance and ML classification score ($r = -0.263$, $p = 0.048$) suggesting its possible role in CRT response prediction. This was a bit surprising, as no difference in this feature was found between the responders and non-responders (see **Table 2**), as well as no correlation with LV EF improvement in our patient cohort ($r < 0.25$, $p > 0.05$). However, selection of this distance as an important feature for ML classifier is in line with clinical studies, where the LAT zone was considered as a target area for LV lead deployment (Chumarnaya et al., 2017; Stephansen et al., 2018; Zubarev et al., 2019; Lahiri et al., 2020). In particular, consistent with clinical data, our results indicate that optimal electrode deployment should be guided by a kind of minimum-maximum optimization with respect to the distances from LAT and disease-induced remodeling area, respectively. Preoperative model-based prediction of such optimal pacing site location seems extremely valuable.

Although ventricular mechanical dyssynchrony was considered with respect to CRT improvements (Duckett et al., 2011; Heydari et al., 2012; Stankovic et al., 2014; Chumarnaya et al., 2017), we did not use mechanical dyssynchrony indices in developing our classifiers because not every patient had these features indicated in the retrospective dataset. We did not find a correlation between the ML response scores generated from the selected hybrid data and the mechanical dyssynchrony indices measured in 34 patients at the baseline ($r < 0.25$, $p > 0.05$ for IVD, T_{max}-T_{min}, SD12). This was not consistent with a correlation between the IVD index and postoperative Δ EF in the patient cohort ($r = 0.32$, $p = 0.029$), and a significant difference in the average IDV indices between responders and nonresponders defined by LV EF improvement (75 ± 17 vs. 63 ± 19 , $p = 0.013$, see **Table 2**). These controversial findings did not allow us to disprove the possible importance of mechanical dyssynchrony indices for ML response prediction, and this hypothesis should be further evaluated on a dataset of bigger size.

It is especially remarkable that each classification model included simulated characteristics of myocardial activation and ECG from the personalized electrophysiology models under LBBB and BiV pacing selected among the most important features. Our best SVM classifier for the EF10 response definition selected three simulated features TAT/MTV under LBBB and BiV pacing, and QRSd under BIV pacing among the 8 most important ones for EF improvement prediction (see **Figure 4**). In particular, two of the three features TAT/MTV and QRSd under BIV pacing correlated with EF improvement ($r = 0.27$ and $r = -0.31$, $p < 0.05$, see **Supplementary Figure S7**), supporting their importance for ML predictions. Note, the *in-silico* indices of electrical dyssynchrony assessed in our study were not selected as important for ML classifiers. These indices were previously suggested by Villongco et al. (2016), who demonstrated a correlation between the post-operational ESV

reduction and the change in the mAT_{STLV} index of inter-ventricular dyssynchrony under BiV pacing against the LBBB baseline on data from 8 patients. In study by Lumens et al. (2015), a combination of clinical data and personalized models of cardiac mechanics and hemodynamics also demonstrated significant role of inter-ventricular electrical dyssynchrony in predicting CRT response defined by an improved LV hemodynamic performance assessed via increase in the maximal derivative of LV pressure (dP/dt_{max}). In contrast, we found no significant correlations between any of the simulated indices of electrical dyssynchrony and echocardiographic CRT response in our cohort ($r < 0.25$, $p > 0.05$). The role of such simulated indices needs further analysis to be performed on a dataset of bigger size.

4.3. Hybrid Dataset Size and Cross-Validation

It is noteworthy that the ML classifiers we developed to predict LV EF improvement can be considered as powerful, especially taking into account the database size of less than 60 entries. In several studies, the ROC AUC was shown to improve significantly with increasing the dataset size from tens to thousands of entries (Feeny et al., 2019). These results allow us to expect further substantial improvement of the quality of the ML classifiers with further increasing the training dataset size. Poor reproducibility of ML results is known as a frequent problem with classifiers developed on small samples. In our case, the restrictive size of the dataset did not allow us to divide data into a conventional 80% training sub-set and 20% testing sub-set, so we had to use 57 Leave-One-Out combinations of data for classifier training.

To confirm the good quality of our classifiers, we tested also a widely-used repeated stratified five-fold cross-validation method with over 1,000 iterations. In this approach we chose 1,000 combinations of 45 training samples from our dataset to train classifiers and the rest 12 samples to test the models. The statistics of the ROC AUC for the five-fold cross-validation approach is shown in **Table 3** (right column) in comparison with that of Leave-One-Out cross-validation (left column). It is seen that average ROC AUCs at five-fold cross-validation are slightly lower than those generated with the Leave-One-Out approach but the latter values fall in the confidence interval of ROC AUC distributions shown by five-fold cross-validation on our dataset. Results demonstrate stability of the ML classifiers we built on our hybrid dataset and confirm the robustness of the ML predictions.

When developing our classifiers, we noticed that the list of features in the cross-validation loop was not steady. This was mainly due to the small sample size. However, even taking this factor into account, we obtained a high accuracy of the constructed classifiers. Testing the SVM classifier for EF10 with a smaller number of features, we found that even for 5 features, the classifier showed the same accuracy as for 8 features (for other criteria of CRT response we observed a lower accuracy with reduced input data dimension). This suggests that with any further increase in the dimension, the classifier cannot converge to optimal solution. Therefore, adding more features to such a small size dataset does not make classifiers more accurate. We

hope that with an increase in the size of dataset, the accuracy of the classifiers will additionally increase due to a more stable feature selection.

4.4. Classifiers for Various Definitions of CRT Response

We used different CRT response definitions to build ML classifiers for our hybrid dataset. Unfortunately, no consensus has been achieved on how to define “response” to CRT Foley et al. (2009), making it difficult to compare different clinical trials and modeling studies. CRT response definition by markers of LV reverse remodeling following device implantation is widely used, and a more than 15% reduction in LV end-systolic volume ($\Delta\text{ESV} < -15\%$) is the most widely accepted criterion (Park et al., 2012). In consistency with that, an optimal cutoff value for ΔESV was defined at 13.5% (sensitivity = 0.719, specificity = 0.719) for a 1-Year hierarchical clinical composite end point in patients who underwent CRT (Uhm et al., 2019). Our earlier 278 patients’ study by Chumarnaya et al. (2021) revealed a 9% cutoff value for ESV reduction for responders. Surprisingly, in our patient cohort, the grouping by either 10% or 15% cutoff for ESV reduction for responders was the same. Therefore, we used the latter definition (ESV15) to determine a positive response to CRT.

A summary of the statistics for the hybrid dataset labeled according to the ESV15 definition of CRT response is presented in **Supplementary Table S3**. ML classifiers with leave-one-out cross-validation on the hybrid data showed a high performance with best ROC AUC of 0.74 (see **Supplementary Figure S9**), and an accuracy of 0.70, sensitivity of 0.87, specificity of 0.37, ppv of 0.73 and npv of 0.58. See also the classifier characteristics for five-fold cross-validation in **Supplementary Table S5**. The ML scores generated by the best classifier built on the ESV15 criterion correlated with post-operational reduction in ESV ($r = -0.27$, $p = 0.039$, see **Supplementary Figure S10**).

The results are slightly less powerful as compared with classifiers built on the EF10 criterion. The latter showed higher ROC AUCs, similar sensitivity, but higher specificity as compared to ESV15 (see **Figure 4** and **Supplementary Figure S9**, **Table 3** and **Supplementary Table S5**). The ML scores based on ESV15 labeling are higher as compared with EF10 scores (0.69 ± 0.18 vs. 0.40 ± 0.35 , $p < 0.01$, respectively) tending to overestimate predictions for the negative response. Note that for both CRT response criteria the average scores are significantly higher in responders vs. nonresponders, indicating good predictive quality of the ML classifiers.

Surprisingly, the sub-sets of 8 most important features selected for classifiers on different response criteria almost did not intersect. For the ESV15 criterion, the pre-operative EDV_{LBBB} showed the primary importance among other inputs in consistency with its correlation with $\Delta\text{ESV}_{\text{CRT}}$ ($r = -0.36$, $p < 0.05$). Another clinical feature selected for classification was IHD/DCM index reflecting the etiology of CHF in patients (see **Supplementary Figure S9**). The rest of the selected features were indices derived from CT/MRI data and simulated features in LBBB and BiV modes of activation. Similar to EF10, the distance from the LV pacing site to infarct/fibrosis zone was the third

in the feature importance range, and simulated TAT/ MTV_{LBBB} was selected for both ESV15 and EF10 criteria together with other model-derived features different between the criteria (see **Figure 4** and **Supplementary Figure S9**).

Unexpectedly, we were not able to generate a predictive model for ESV15 criterion from the clinical feature subset suggested in Feeny et al. (2019) with ROC AUC > 0.5 on the dataset for our patient cohort. When we calculated ML scores using the calculator from Feeny et al. (2019) for our responders and nonresponders defined by ESV15 criterion, the average ML scores did not differ between the groups, while the ML scores based on the hybrid data were significantly different (see **Supplementary Table S3**). These findings also point to the power of model-driven data in CRT response prediction.

We also compared the accuracy of ML classifiers built on the hybrid dataset for CRT response defined by 5, 10, 15% LV EF improvement and by coupled EF10 and ESV15 criteria (see **Supplementary Tables S4, S5** in the Materials). For every response definition, our best classifiers demonstrate improved performance as compared with all clinical and ML predictors reported in Feeny et al. (2019). Note again that our hybrid data classifiers were trained on a dataset of much smaller size than previously published (Feeny et al., 2019). Like in Feeny et al. (2019), for different ΔEF criteria an average accuracy of the predictive models increases with the cutoff for the LV EF improvement for CRT responders. However, the sensitivity and predictive positive value of the models tend to decrease with increasing the ΔEF cutoff, while both the specificity and predictive negative value increase. Thus, ML scores tend to underestimate the probability of a super-response. The best balance between sensitivity and specificity was shown for the $\Delta\text{LV EF} > 10\%$ definition of CRT response which also demonstrates the best ROC AUC among other criteria, thus supporting the choice of this criterion for response evaluation in patients.

4.5. Principal Coordinate Analysis and Unsupervised ML Clustering for CRT Response Prediction

The supervised multi-variable classifiers we discussed in the previous sections were built using feature selection approaches where input values are intrinsic functional characteristics of the processes. Often, in ML algorithms principal component analysis (PCA) is used for data dimension reduction, which allows more objective exclusion of collinearity between the input features. We tested the PCA in combination with Logistic Regression using different numbers of PCs for classifier development. We evaluated ROC AUCs using from 2 to 10 PCs, and obtained the best ROC AUC of 0.70 for 5 PCs (**Supplementary Figure S8**, left panel) with explainable variance of 0.58. This ROC AUC is much lower than the best values demonstrated by other ML classifiers we developed using row feature values.

As we showed in the previous section, classification results depend on the positive CRT response definition used for data labeling. Another ML approach is unsupervised ML

data clustering based on their similarity without the help of class labels. We performed clustering by K-means, used in recent studies for CRT response evaluation (Cikes et al., 2019; Feeny et al., 2020). Using K-means clustering on the two first PCs, we differentiated our dataset into 2 clusters (see **Supplementary Figure S8**, right panel). However, mean ΔEF_{CRT} were not significantly different between the groups ($8.0 \pm 8.6\%$ vs. $10.7 \pm 8.3\%$, $p = 0.27$), similar to no difference in mean ΔESV_{CRT} ($-20 \pm 38\%$ vs. $-34 \pm 31\%$, $p = 0.15$). Moreover, distribution between the two clusters of CRT responders and non-responders defined by the EF10 criterion shows random assignment to the groups (see **Supplementary Figure S8**, right panel). These findings suggest that unsupervised learning on a small dataset does not allow one to reliably differentiate pre-operational data into groups clearly associated with CRT response characteristics. In contrast, the supervised ML algorithms we developed provided valuable predictions of CRT response showing the potential of model-derived features.

5. LIMITATIONS

There are several limitations in our study that have to be overcome to make our approach actually usable in clinic. First, ventricular geometry in our personalized models was derived from CT images obtained after CRT device implantation, not before it. This was essential for this proof-of-concept study because it allowed us to define the precise location of pacing electrodes and to fit our models to both LBBB and BiV ECG data for the same ventricular geometry thus demonstrating the potential of our models for reproducing real clinical data. Despite supposed difference in the ventricular geometry our simulated ECGs in the LBBB mode had a high correlation with pre-operative clinical ECGs ($r = 0.84$, $p < 0.05$), thus demonstrating the effect of ventricular geometry as being secondary. Of course, the reverse remodeling of the ventricles after CRT may affect the difference in model simulations before and after operation. That is why we primarily focused on the CRT response definition based on the EF improvement which has low-to-moderate correlation with ventricular remodeling in our patient cohort. The main idea of using model-derived biomarkers for CRT response prediction was the possibility to assess the primary effect of ventricular synchronization itself on the electrophysiological characteristics of activation, where changes in the geometry seem less important.

The second limitation is that we used here a simplified Eikonal equation allowing us to reproduce the QRS complex of an ECG but not an entire ECG signal. Moreover, we used a simplified approach to tailor the model to personalized data focusing on the mean QRSd from 12-lead ECG as a target for the parameter identification problem. Then we used the maximal QRSd as a model biomarker for building a classifier. The QRS morphology may provide much more information for tailoring personalized electrophysiology models and then for CRT response predictions. A recent study of Camps et al. (2021) showed a way toward more accurate personalization

of the activation processes in ventricles based on the QRS signals recorded in patients, which may be useful for model improvement. Feeny et al. (2020) also demonstrated the power of the entire ECG signal for ML predictions of CRT response, suggesting that the use of the entire simulated ECG under BiV pacing may further improve ML predictors. In future studies, we will use more adequate mono-domain models to reproduce both activation and repolarisation of myocardium, and will assess the contribution of entire ECG signals to the accuracy of CRT predictive models.

Next, we have shown high importance of the distance from the LV pacing site to the myocardial damage area in ML predictions. In this study, we did not have access to raw MRI data from patients to be able to derive accurate information on post-infarction scar or fibrosis morphology. We used only textual descriptions of the infarct zone location with a segment accuracy within a 17-segment AHA LV model from an expert who evaluated MRI data in patients. There are great examples of using detailed morphology of myocardial remodeling area in personalized cardiac models for predicting the risk of cardiac arrhythmia and patient stratification (Lopez-Perez et al., 2019). We think more objective information on the scar and fibrosis morphology may improve predictive models of CRT response as well.

In this study, for CRT response prediction we used simulated characteristics of ventricular activation and ECG derived from electrophysiological models. The use of the model features characterizing ventricular excitation is justified by the essence of the therapy, which ensures electrical synchronization of ventricular activation, and the success of this synchronization determines the outcome of the operation. However, the goal of CRT implantation is the synchronization of ventricular contraction and subsequent improvement in the mechanical performance of the ventricles. This opens up a further direction for studies using electromechanical models of cardiac activity which are being developed in modeling community including our group (Sugiura et al., 2012; Chabiniok et al., 2016; Isotani et al., 2020) and which are able to predict directly EF, dP/dt_{max} changes and other mechanical biomarkers of CRT response. Such models were already used for clinical data analysis in CRT patients by several groups (Sermesant et al., 2012; Okada et al., 2017; Lee et al., 2018; Isotani et al., 2020), demonstrating the power of such simulations for CRT response predictions. In particular, we believe that reduced mechanical models using regression or ML approaches to reproduce the behavior of complex 3D models such as developed with our participation (Di Achille et al., 2018) would be the best choice in terms of possible clinical application of model simulations.

In this study, the precise RV/LV pacing lead location was determined from the post-operative CT scans for the patients. Thus, the same pacing sites were used in our BiV model simulations to exclude the effects of uncertainty in lead position on the ML prediction results. This was the first step in the validation of our new technique suggesting its high potential in CRT response prediction. In real practice, however, patient selection should be done before the clinical procedure. The main advantage of using personalized computational models

is the possibility to compute characteristics of ventricular activation from any accessible pacing sites. Indeed, if the coronary sinus anatomy is appropriate (which is possible to derive from CT data), one can predict an accessible area for pacing electrode installation. Hence, this area could be included in a personalized ventricular model to simulate BiV pacing. Moreover, RV/LV electrode locations can be varied throughout the ventricle surfaces. Thus, simulations performed prior to clinical intervention can be used to directly assess the effects of BiV pacing on ventricular activation time, ECG biomarkers and electrical dyssynchrony indices from various pacing sites. Also, it might be helpful during selection the best possible electrode location optimizing ventricular synchronization (or any other optimized function) from model simulations. To sum up, the next step of our approach development is the prediction of pacing lead location prior to interventions by using features extracted from the results of the simulations of BiV pacing.

A number of clinical and simulation studies paid great attention to the possibility to target pacing lead implantation (Bakos et al., 2014; Nguyễn et al., 2018b; Sieniewicz et al., 2018). Different criteria for optimal electrode location were discussed in the literature, including optimization of electrical synchronization characteristics, e.g., maximal narrowing of QRSd, or minimization of interventricular dyssynchrony; maximal proximity to the late activation zone or late contraction zone; avoiding the match with infarct zone; or maximizing the mechanical performance characteristics, e.g., dP/dt_{max} (Bakos et al., 2014; Nguyễn et al., 2018a; Isotani et al., 2020; Albatat et al., 2021). The use of simulations from personalized models in CRT response prediction opens the possibility to re-evaluate these hypotheses and suggest a new strategy for implantation planning with allowing for model-based prediction of optimal location for pacing electrodes. We are going to test this hypothesis in a future prospective study.

The long-term goal of CRT is to reduce morbidity and mortality in heart failure patients with reduced left ventricular function and intraventricular conduction delay. Several studies tested ML approaches for predicting outcomes after CRT in terms of patient survival and frequency of adverse events in the longer term after operation (Kalscheur et al., 2018; Tokodi et al., 2020). We had no sufficient data to perform such analysis using simulated data, and this could be another new direction of future studies.

Last but not least, in this study we had a limited data sample from 57 patients. Of course, this number is quite small for ML algorithms operating on thousands of entries with a possibility to use separate subsets for training and testing. However, our predictive models based on hybrid data from clinic and computational models of cardiac activity have demonstrated high performance with accuracy much higher than that demonstrated by predictors developed on clinical data from a thousand of patients. We used feature selection within the cross-validation loop to eliminate any bias factors. The performance of the best classifier was also higher than that of available classifiers based on clinical data (which are based on much larger datasets). In addition, in this study we used simple models that do not tend to overfit on small datasets. Also, we didn't do any

hyperparameter search, as a result of which the models could be overfit. We believe that feature selection in the cross-validation loop will be more stable on a larger dataset. This inspires hope that a larger dataset and more informative data from time-dependent simulated signals may further improve CRT response predictions.

6. CONCLUSIONS

We have developed a new technology combining personalized heart modeling and supervised ML techniques to predict CHF patient improvement under CRT. We constructed 57 multimodal image-based personalized models of ventricular geometry and myocardial damage area. The models were used to simulate ventricular activation and ECG on the patient torso at LBBB and BiV pacing. Supervised ML algorithms used features extracted from the results of the simulations combined with additional clinical indices and MRI/CT derived features.

Despite a limited dataset, we have developed several high-performing ML classifiers from the hybrid dataset. The best SVM classifier showed an accuracy of 0.82, sensitivity of 0.85, and specificity of 0.78. The classifier on hybrid data outperformed ML predictors built on clinical data only.

The majority of the most relevant features selected from the hybrid dataset for the ML classifiers were model-driven indices, suggesting their great power for CRT response prediction. Distance from the LV pacing site to the infarct/fibrosis area and features extracted from simulations under BiV pacing were shown as the most important features for patient classification.

The novel proposed approach has great potential clinical implications suggesting patient care improvement. With an ML classifier on hybrid data created and thoroughly validated, one would be to assess with a high degree of accuracy the likelihood of improvement in a particular patient's condition prior to a CRT procedure. In this way, ML scores would be computed for the patient using personalized model simulations for BIV pacing (or other type of stimulation) from various accessible pacing site locations. The range of generated ML scores would classify this patient as a potential responder or nonresponder to the therapy, thus supporting individual selection for it. At the same time, the best pacing site location predicted from the model simulation results and corresponding ML scores could further be used to guide electrode deployment during CRT procedure optimizing the patient output. This technology would be especially effective in the merging of detailed multimodal imaging data on the ventricular geometry and structure of myocardial damage (infarct, fibrosis, inflammation, adipose), coronary sinus anatomy, His-Purkinje conduction system and information on cellular remodeling in the myocardial tissue.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Ethical Committee of Almazov National Medical Research Centre. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

OS, SK, TC, AD, and AB participated in conceptualization and designing of the study, data collection, and preprocessing. SK implemented the Eikonal and pseudobidomain models, implemented machine learning models and analyzed the data, interpreted the results and was a major contributor to performed study. SZ, AB, and AD contributed in CT images segmentation and anatomical models construction. AD developed Purkinje system model and performed computational simulations for LBBB. AB was a major contributor to the anatomical model preprocessing and enhancement, infarction assignment, and performed computational simulations for BiV pacing. SZ, TL,

VL, DL, and TC participated in clinical data collection and preprocessing. TC contributed in clinical data analysis, result interpretation, providing descriptive statistics. VG contributed in designing of the machine learning study. OS made a major contribution to the conceptualization and designing of the study, supervised all stages of study execution and was a major contributor to writing the manuscript. All authors contributed to manuscript preparation and approved the final version of the manuscript.

FUNDING

This work was supported by Russian Science Foundation grant no. 19-14-00134.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2021.753282/full#supplementary-material>

REFERENCES

- Abraham, W. T., Fisher, W. G., Smith, A. L., Delurgio, D. B., Leon, A. R., Loh, E., et al. (2002). Cardiac resynchronization in chronic heart failure. *N. Engl. J. Med.* 346, 1845–1853. doi: 10.1056/NEJMoa013168
- Albatat, M., Arevalo, H., Bergsland, J., Strøm, V., Balasingham, I., and Odland, H. H. (2021). Optimal pacing sites in cardiac resynchronization by left ventricular activation front analysis. *Comput. Biol. Med.* 128:104159. doi: 10.1016/j.compbiomed.2020.104159
- Aronis, K. N., Prakosa, A., Bergamaschi, T., Berger, R. D., Boyle, P. M., Chrispin, J., et al. (2021). Characterization of the electrophysiologic remodeling of patients with ischemic cardiomyopathy by clinical measurements and computer simulations coupled with machine learning. *Front. Physiol.* 12:684149. doi: 10.3389/fphys.2021.684149
- Auricchio, A., and Prinzen, F. W. (2017). Enhancing response in the cardiac resynchronization therapy patient: the 3B perspective—bench, bits, and bedside. *JACC Clin. Electrophysiol.* 3, 1203–1219. doi: 10.1016/j.jacep.2017.08.005
- Bakos, Z., Markstad, H., Ostefeld, E., Carlsson, M., Roijer, A., and Borgquist, R. (2014). Combined preoperative information using a bullseye plot from speckle tracking echocardiography, cardiac CT scan, and MRI scan: Targeted left ventricular lead implantation in patients receiving cardiac resynchronization therapy. *Eur. Heart J. Cardiovasc. Imaging* 15, 523–531. doi: 10.1093/ehjci/jet233
- Bayer, J. D., Blake, R. C., Plank, G., and Trayanova, N. A. (2012). A novel rule-based algorithm for assigning myocardial fiber orientation to computational heart models. *Ann. Biomed. Eng.* 40, 2243–2254. doi: 10.1007/s10439-012-0593-5
- Bishop, M. J., and Plank, G. (2011). Bidomain ECG simulations using an augmented monodomain model for the cardiac source. *IEEE Trans. Biomed. Eng.* 58, 2297–2307. doi: 10.1109/TBME.2011.2148718
- Bristow, M. R., Saxon, L. A., Boehmer, J., Krueger, S., Kass, D. A., De Marco, T., et al. (2004). Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N. Engl. J. Med.* 350, 2140–2150. doi: 10.1056/NEJMoa032423
- Camps, J., Lawson, B., Drovandi, C., Mincholé, A., Wang, Z. J., Grau, V., et al. (2021). Inference of ventricular activation properties from non-invasive electrocardiography. *Med. Image Anal.* 73:102143. doi: 10.1016/j.media.2021.102143
- Cerqueira, M. D., Weissman, N. J., Dilsizian, V., Jacobs, A. K., Kaul, S., Laskey, W. K., et al. (2002). Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the American Heart Association. *Circulation* 105, 539–542. doi: 10.1161/hc0402.102975
- Chabiniok, R., Wang, V. Y., Hadjicharalambous, M., Asner, L., Lee, J., Sermesant, M., et al. (2016). Multiphysics and multiscale modelling, data-model fusion and integration of organ physiology in the clinic: ventricular cardiac mechanics. *Interface Focus* 6:20150083. doi: 10.1098/rsfs.2015.0083
- Chalil, S., Foley, P. W., Muhyaldeen, S. A., Patel, K. C., Yousef, Z. R., Smith, R. E., et al. (2007). Late gadolinium enhancement-cardiovascular magnetic resonance as a predictor of response to cardiac resynchronization therapy in patients with ischaemic cardiomyopathy. *Europace* 9, 1031–1037. doi: 10.1093/europace/eum133
- Chumarnaya, T., Lyubimtseva, T., Solodushkin, S., Lebedeva, V., Lebedev, D., and Solovyova, O. (2021). Evaluation of the effectiveness of cardiac resynchronization therapy in the long-term postoperative period. *Russian J. Cardiol. Accepted*. doi: 10.15829/1560-4071-2021-4531
- Chumarnaya, T., Trifanova, M., Lyubimtseva, T., Lebedeva, V., Poroshin, I., Trukshina, M., et al. (2017). Impact of inter-ventricular lead distance on cardiac resynchronization therapy outcomes. *Comput. Cardiol.* 44, 1–4. doi: 10.22489/CinC.2017.290-106
- Cikes, M., Sanchez-Martinez, S., Claggett, B., Duchateau, N., Piella, G., Butakoff, C., et al. (2019). Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur. J. Heart Fail.* 21, 74–85. doi: 10.1002/ejhf.1333
- Di Achille, P., Harouni, A., Khamzin, S., Solovyova, O., Rice, J. J., and Gurev, V. (2018). Gaussian process regressions for inverse problems and parameter searches in models of ventricular mechanics. *Front. Physiol.* 9:1002. doi: 10.3389/fphys.2018.01002
- Dobrzynski, H., Anderson, R. H., Atkinson, A., Borbas, Z., D'Souza, A., Fraser, J. F., et al. (2013). Structure, function and clinical relevance of the cardiac conduction system, including the atrioventricular ring and outflow tract tissues. *Pharmacol. Ther.* 139, 260–288. doi: 10.1016/j.pharmthera.2013.04.010
- Duckett, S. G., Ginks, M., Shetty, A. K., Bostock, J., Gill, J. S., Hamid, S., et al. (2011). Invasive acute hemodynamic response to guide left ventricular lead implantation predicts chronic remodeling in patients undergoing cardiac resynchronization therapy. *J. Am. Coll. Cardiol.* 58, 1128–1136. doi: 10.1016/j.jacc.2011.04.042

- Feeny, A. K., Rickard, J., Patel, D., Toro, S., Trulock, K. M., Park, C. J., et al. (2019). Machine learning prediction of response to cardiac resynchronization therapy. *Circ. Arrhythm. Electrophysiol.* 12:e007316. doi: 10.1161/CIRCEP.119.007316
- Feeny, A. K., Rickard, J., Trulock, K. M., Patel, D., Toro, S., Moennich, L. A., et al. (2020). Machine learning of 12-lead qrs waveforms to identify cardiac resynchronization therapy patients with differential outcomes. *Circ. Arrhythm. Electrophysiol.* 13:e008210. doi: 10.1161/CIRCEP.119.008210
- Foley, P. W., Leyva, F., and Frenneaux, M. P. (2009). What is treatment success in cardiac resynchronization therapy? *Europace* 11(Suppl. 5):v58–v65. doi: 10.1093/europace/eup308
- Franzone, P. C., and Guerri, L. (1993). Spreading of excitation in 3-d models of the anisotropic cardiac tissue. i. validation of the eikonal model. *Math. Biosci.* 113, 145–209. doi: 10.1016/0025-5564(93)90001-Q
- Heijman, J., Sutanto, H., Crijns, H. J. G. M., Nattel, S., and Trayanova, N. A. (2021). Computational models of atrial fibrillation: achievements, challenges, and perspectives for improving clinical care. *Cardiovasc. Res.* 117, 1682–1699. doi: 10.1093/cvr/cvab138
- Heydari, B., Jerosch-Herold, M., and Kwong, R. Y. (2012). Imaging for planning of cardiac resynchronization therapy. *JACC Cardiovasc. Imaging* 5, 93–110. doi: 10.1016/j.jcmg.2011.11.006
- Hsu, J. C., Solomon, S. D., Bourgoun, M., McNitt, S., Goldenberg, I., Klein, H., et al. (2012). Predictors of super-response to cardiac resynchronization therapy and associated improvement in clinical outcome: the MADIT-CRT (multicenter automatic defibrillator implantation trial with cardiac resynchronization therapy) study. *J. Am. Coll. Cardiol.* 59, 2366–2373. doi: 10.1016/j.jacc.2012.01.065
- Hu, S.-Y., Santus, E., Forsyth, A. W., Malhotra, D., Haimson, J., Chatterjee, N. A., et al. (2019). Can machine learning improve patient selection for cardiac resynchronization therapy? *PLoS ONE* 14:e0222397. doi: 10.1371/journal.pone.0222397
- Huntjens, P. R., Ploux, S., Strik, M., Walmsley, J., Ritter, P., Haissaguerre, M., et al. (2018). Electrical substrates driving response to cardiac resynchronization therapy: a combined clinical-computational evaluation. *Circ. Arrhythm. Electrophysiol.* 11:e005647. doi: 10.1161/CIRCEP.117.005647
- Isotani, A., Yoneda, K., Iwamura, T., Watanabe, M., Ichi Okada, J., Washio, T., et al. (2020). Patient-specific heart simulation can identify non-responders to cardiac resynchronization therapy. *Heart Vessels.* 35, 1135–1147. doi: 10.1007/s00380-020-01577-1
- Kalscheur, M. M., Kipp, R. T., Tattersall, M. C., Mei, C., Buhr, K. A., Demets, D. L., et al. (2018). Machine learning algorithm predicts cardiac resynchronization therapy outcomes: lessons from the COMPANION trial. *Circ. Arrhythm. Electrophysiol.* 11:e005499. doi: 10.1161/CIRCEP.117.005499
- Keener, J. P. (1991). An eikonal-curvature equation for action potential propagation in myocardium. *J. Math. Biol.* 29, 629–651. doi: 10.1007/BF00163916
- Krittawanong, C., Zhang, H., Wang, Z., Aydar, M., and Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *J. Am. Coll. Cardiol.* 69, 2657–2664. doi: 10.1016/j.jacc.2017.03.571
- Lahiri, A., Chahadi, F. K., Ganesan, A. N., and McGavigan, A. D. (2020). Characteristics that predict response after cardiac resynchronization therapy. *Curr. Cardiovasc. Risk Rep.* 14, 1–10. doi: 10.1007/s12170-020-00640-w
- Lee, A., Nguyen, U., Razeghi, O., Gould, J., Sidhu, B., Sieniewicz, B., et al. (2019). A rule-based method for predicting the electrical activation of the heart with cardiac resynchronization therapy from non-invasive clinical data. *Med. Image Anal.* 57, 197–213. doi: 10.1016/j.media.2019.06.017
- Lee, A. W. C., Mendonca Costa, C., Strocchi, M., Rinaldi, C. A., and Niederer, S. A. (2018). Computational modeling for cardiac resynchronization therapy. *J. Cardiovasc. Transl. Res.* 11, 92–108. doi: 10.1007/s12265-017-9779-4
- Logg, A., and Wells, G. N. (2010). Dolfin: automated finite element computing. *ACM Trans. Math. Softw.* 37, 1–18. doi: 10.1145/1731022.1731030
- Lopez-Jimenez, F., Attia, Z., Arruda-Olson, A. M., Carter, R., Chareonthaitawee, P., Jouni, H., et al. (2020). Artificial intelligence in cardiology: present and future. *Mayo Clin. Proc.* 95, 1015–1039. doi: 10.1016/j.mayocp.2020.01.038
- Lopez-Perez, A., Sebastian, R., Izquierdo, M., Ruiz, R., Bishop, M., and Ferrero, J. M. (2019). Personalized cardiac computational models: from clinical data to simulation of infarct-related ventricular tachycardia. *Front. Physiol.* 10:580. doi: 10.3389/fphys.2019.00580
- Lumens, J., Tayal, B., Walmsley, J., Delgado-Montero, A., Huntjens, P. R., Schwartzman, D., et al. (2015). Differentiating electromechanical from non-electrical substrates of mechanical discoordination to identify responders to cardiac resynchronization therapy. *Circ. Cardiovasc. Imaging* 8:e003744. doi: 10.1161/CIRCIMAGING.115.003744
- Marsan, N. A., Westenberg, J. J., Ypenburg, C., van Bommel, R. J., Roes, S., Delgado, V., et al. (2009). Magnetic resonance imaging and response to cardiac resynchronization therapy: relative merits of left ventricular dyssynchrony and scar tissue. *Eur. Heart J.* 30, 2360–2367. doi: 10.1093/eurheartj/ehp280
- Mullens, W., Grimm, R. A., Verga, T., Dresing, T., Starling, R. C., Wilkoff, B. L., et al. (2009). Insights from a cardiac resynchronization optimization clinic as part of a heart failure disease management program. *J. Am. Coll. Cardiol.* 53, 765–773. doi: 10.1016/j.jacc.2008.11.024
- Nguyèn, U. C., Claridge, S., Vernoooy, K., Engels, E. B., Razavi, R., Rinaldi, C. A., et al. (2018a). Relationship between vectorcardiographic QRSarea, myocardial scar quantification, and response to cardiac resynchronization therapy. *J. Electrocardiol.* 51, 457–463. doi: 10.1016/j.jelectrocard.2018.01.009
- Nguyèn, U. C., Cluitmans, M. J., Luermans, J. G., Strik, M., de Vos, C. B., Kietselaer, B. L., et al. (2018b). Visualisation of coronary venous anatomy by computed tomography angiography prior to cardiac resynchronisation therapy implantation. *Neth. Heart J.* 26, 433–444. doi: 10.1007/s12471-018-1132-2
- Okada, J., I., Washio, T., Nakagawa, M., Watanabe, M., Kadooka, Y., et al. (2017). Multi-scale, tailor-made heart simulation can predict the effect of cardiac resynchronization therapy. *J. Mol. Cell. Cardiol.* 108, 17–23. doi: 10.1016/j.yjmcc.2017.05.006
- Ono, N., Yamaguchi, T., Ishikawa, H., Arakawa, M., Takahashi, N., Saikawa, T., et al. (2009). Morphological varieties of the Purkinje fiber network in mammalian hearts, as revealed by light and electron microscopy. *Arch. Histol. Cytol.* 72, 139–149. doi: 10.1679/aohc.72.139
- Park, M. Y., Altman, R. K., Orencole, M., Kumar, P., Parks, K. A., Heist, K. E., et al. (2012). Characteristics of responders to cardiac resynchronization therapy: the impact of echocardiographic left ventricular volume. *Clin. Cardiol.* 35, 779–780. doi: 10.1002/clc.22043
- Pezel, T., Milka, D., Logeart, D., Cohen-Solal, A., Beauvais, F., Henry, P., et al. (2021). Characterization of non-response to cardiac resynchronization therapy by post-procedural computed tomography. *Pacing Clin. Electrophysiol.* 44, 135–144. doi: 10.1111/pace.14134
- Pezzuto, S., Kal'avský, P., Potse, M., Prinzen, F. W., Auricchio, A., and Krause, R. (2017). Evaluation of a rapid anisotropic model for ecg simulation. *Front. Physiol.* 8:265. doi: 10.3389/fphys.2017.00265
- Potse, M., Krause, D., Kroon, W., Murzilli, R., Muzzarelli, S., Regoli, F., et al. (2014). Patient-specific modelling of cardiac electrophysiology in heart-failure patients. *Europace* 16(Suppl. 4):iv56–iv61. doi: 10.1093/europace/euu257
- Pullan, A. J., Tomlinson, K. A., and Hunter, P. J. (2006). A finite element method for an eikonal equation model of myocardial excitation wavefront propagation. *SIAM J. Appl. Math.* 63, 324–350. doi: 10.1137/S0036139901389513
- Sahli Costabal, F., Hurtado, D. E., and Kuhl, E. (2016). Generating Purkinje networks in the human heart. *J. Biomech.* 49, 2455–2465. doi: 10.1016/j.jbiomech.2015.12.025
- Sermesant, M., Chabiniok, R., Chinchapatnam, P., Mansi, T., Billet, F., Moireau, P., et al. (2012). Patient-specific electromechanical models of the heart for the prediction of pacing acute effects in CRT: a preliminary clinical validation. *Med. Image Anal.* 16, 201–215. doi: 10.1016/j.media.2011.07.003
- Shade, J. K., Cartoski, M. J., Nikolov, P., Prakosa, A., Doshi, A., Binka, E., et al. (2020). Ventricular arrhythmia risk prediction in repaired tetralogy of fallot using personalized computational cardiac models. *Heart Rhythm.* 17, 408–414. doi: 10.1016/j.hrthm.2019.10.002
- Shade, J. K., Prakosa, A., Popescu, D. M., Yu, R., Okada, D. R., Chrispin, J., et al. (2021). Predicting risk of sudden cardiac death in patients with cardiac sarcoidosis using multimodality imaging and personalized heart modeling in a multivariable classifier. *Sci. Adv.* 7, 8020–8048. doi: 10.1126/sciadv.abi8020
- Sieniewicz, B. J., Gould, J., Porter, B., Sidhu, B. S., Behar, J. M., Claridge, S., et al. (2018). Optimal site selection and image fusion guidance technology to facilitate cardiac resynchronization therapy. *Expert. Rev. Med. Devices* 15, 555–570. doi: 10.1080/17434440.2018.1502084
- Stankovic, I., Aaronson, M., Smith, H. J., Vörös, G., Kongsgaard, E., Neskovic, A. N., et al. (2014). Dynamic relationship of left-ventricular dyssynchrony and

- contractile reserve in patients undergoing cardiac resynchronization therapy. *Eur. Heart J.* 35, 48–55. doi: 10.1093/eurheartj/ehz294
- Stephansen, C., Sommer, A., Kronborg, M. B., Jensen, J. M., Bouchelouche, K., and Nielsen, J. C. (2018). Electrically guided versus imaging-guided implant of the left ventricular lead in cardiac resynchronization therapy: a study protocol for a double-blinded randomized controlled clinical trial (ElectroCRT). *Trials* 19:600. doi: 10.1186/s13063-018-2930-y
- Sugiura, S., Washio, T., Hatano, A., Okada, J., Watanabe, H., and Hisada, T. (2012). Multi-scale simulations of cardiac electrophysiology and mechanics using the University of Tokyo heart simulator. *Prog. Biophys. Mol. Biol.* 110, 380–389. doi: 10.1016/j.pbiomolbio.2012.07.001
- Sung, E., Prakosa, A., Aronis, K. N., Zhou, S., Zimmerman, S. L., Tandri, H., et al. (2020). Personalized digital-heart technology for ventricular tachycardia ablation targeting in hearts with infiltrating adiposity. *Circ. Arrhythm. Electrophysiol.* 13:e008912. doi: 10.1161/CIRCEP.120.008912
- ten Tusscher, K. H. W. J. (2006). Alternans and spiral breakup in a human ventricular tissue model. *AJPHeart Circ. Physiol.* 291, H1088–H1100. doi: 10.1152/ajpheart.00109.2006
- Tokodi, M., Behon, A., Merkel, E. D., Kovács, A., Tösér, Z., Sárkány, A., et al. (2021). Sex-specific patterns of mortality predictors among patients undergoing cardiac resynchronization therapy: a machine learning approach. *Front. Cardiovasc. Med.* 8:87. doi: 10.3389/fcvm.2021.611055
- Tokodi, M., Schwertner, W. R., Kovács, A., Tösér, Z., Staub, L., Sárkány, A., et al. (2020). Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. *Eur. Heart J.* 41, 1747–1756. doi: 10.1093/eurheartj/ehz902
- Uhm, J. S., Oh, J., Cho, I. J., Park, M., Kim, I. S., Jin, M. N., et al. (2019). Left ventricular end-systolic volume can predict 1-year hierarchical clinical composite end point in patients with cardiac resynchronization therapy. *Yonsei Med. J.* 60, 48–55. doi: 10.3349/ymj.2019.60.1.48
- Vernooy, K., van Deursen, C. J. M., Strik, M., and Prinzen, F. W. (2014). Strategies to improve cardiac resynchronization therapy. *Nat. Rev. Cardiol.* 11, 481–493. doi: 10.1038/nrcardio.2014.67
- Villongco, C. T., Krummen, D. E., Omens, J. H., and McCulloch, A. D. (2016). Non-invasive, model-based measures of ventricular electrical dyssynchrony for predicting CRT outcomes. *Europace* 18(Suppl_4):iv104–iv112. doi: 10.1093/europace/euw356
- Yu, C. M., Bax, J. J., and Gorcsan, J. (2009). Critical appraisal of methods to assess mechanical dyssynchrony. *Curr. Opin. Cardiol.* 24, 18–28. doi: 10.1097/HCO.0b013e32831bc34e
- Zubarev, S., Chmelevsky, M., Potyagaylo, D., Budanova, M., Trukshina, M., Rud, S., et al. (2019). “Noninvasive electrocardiographic imaging with magnetic resonance tomography in candidates for cardiac resynchronization therapy,” in *Computing in Cardiology, 2019-September* (Singapore: IEEE).

Conflict of Interest: VG was employed by IBM Research.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Khamzin, Dokuchaev, Bazhutina, Chumarnaya, Zubarev, Lyubimtseva, Lebedeva, Lebedev, Gurev and Solovyova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.