

Original Paper

# Disease Progression of Hypertrophic Cardiomyopathy: Modeling Using Machine Learning

Matej Pičulin<sup>1\*</sup>, PhD; Tim Smole<sup>1\*</sup>, MSc; Bojan Žunkovič<sup>1\*</sup>, PhD; Enja Kokalj<sup>1\*</sup>, MSc; Marko Robnik-Šikonja<sup>1\*</sup>, Prof Dr, PhD; Matjaž Kukar<sup>1\*</sup>, Prof Dr, PhD; Dimitrios I Fotiadis<sup>2\*</sup>, Prof Dr, PhD; Vasileios C Pezoulas<sup>2\*</sup>, MSc; Nikolaos S Tachos<sup>2\*</sup>, MSc; Fausto Barlocco<sup>3,4\*</sup>, PhD; Francesco Mazzarotto<sup>3,4,5\*</sup>, PhD; Dejana Popović<sup>6\*</sup>, MD, Prof Dr, PhD; Lars S Maier<sup>7\*</sup>, Prof Dr, PhD; Lazar Velicki<sup>8,9\*</sup>, MD, Prof Dr, PhD; Iacopo Olivetto<sup>3,4\*</sup>, MD, Prof Dr; Guy A MacGowan<sup>10\*</sup>, MD, PhD; Djordje G Jakovljević<sup>10,11\*</sup>, MD, Prof Dr, PhD; Nenad Filipović<sup>12\*</sup>, Prof Dr, PhD; Zoran Bosnić<sup>1\*</sup>, Prof Dr, PhD

<sup>1</sup>Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

<sup>2</sup>Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, Greece

<sup>3</sup>Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy

<sup>4</sup>Cardiomyopathy Unit, Careggi University Hospital, University of Florence, Florence, Italy

<sup>5</sup>National Heart and Lung Institute, Imperial College London, London, United Kingdom

<sup>6</sup>Clinic for Cardiology, Clinical Center of Serbia, University of Belgrade, Belgrade, Serbia

<sup>7</sup>Department of Internal Medicine II (Cardiology, Pneumology, Intensive Care Medicine), University Hospital Regensburg, Regensburg, Germany

<sup>8</sup>Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia

<sup>9</sup>Institute of Cardiovascular Diseases Vojvodina, Sremska Kamenica, Serbia

<sup>10</sup>Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

<sup>11</sup>Faculty of Health and Life Sciences, Coventry University, Coventry, United Kingdom

<sup>12</sup>Bioengineering Research and Development Center, Kragujevac, Serbia

\* all authors contributed equally

## Corresponding Author:

Matej Pičulin, PhD

Faculty of Computer and Information Science

University of Ljubljana

Večna pot 113

Ljubljana, 1000

Slovenia

Phone: 386 1 479 8226

Email: [matej.piculin@fri.uni-lj.si](mailto:matej.piculin@fri.uni-lj.si)

## Abstract

**Background:** Cardiovascular disorders in general are responsible for 30% of deaths worldwide. Among them, hypertrophic cardiomyopathy (HCM) is a genetic cardiac disease that is present in about 1 of 500 young adults and can cause sudden cardiac death (SCD).

**Objective:** Although the current state-of-the-art methods model the risk of SCD for patients, to the best of our knowledge, no methods are available for modeling the patient's clinical status up to 10 years ahead. In this paper, we propose a novel machine learning (ML)-based tool for predicting disease progression for patients diagnosed with HCM in terms of adverse remodeling of the heart during a 10-year period.

**Methods:** The method consisted of 6 predictive regression models that independently predict future values of 6 clinical characteristics: left atrial size, left atrial volume, left ventricular ejection fraction, New York Heart Association functional classification, left ventricular internal diastolic diameter, and left ventricular internal systolic diameter. We supplemented each prediction with the explanation that is generated using the Shapley additive explanation method.

**Results:** The final experiments showed that predictive error is lower on 5 of the 6 constructed models in comparison to experts (on average, by 0.34) or a consortium of experts (on average, by 0.22). The experiments revealed that semisupervised learning

and the artificial data from virtual patients help improve predictive accuracies. The best-performing random forest model improved  $R^2$  from 0.3 to 0.6.

**Conclusions:** By engaging medical experts to provide interpretation and validation of the results, we determined the models' favorable performance compared to the performance of experts for 5 of 6 targets.

(*JMIR Med Inform* 2022;10(2):e30483) doi: [10.2196/30483](https://doi.org/10.2196/30483)

## KEYWORDS

hypertrophic cardiomyopathy; disease progression; machine learning; artificial intelligence; AI; ML; cardiomyopathy; cardiovascular disease; sudden cardiac death; SCD; prediction; prediction model; validation

## Introduction

### Background

Recent reviews of machine learning (ML) applications in cardiovascular medicine [1,2] suggest that the use of ML is on the rise and that it is being adopted by doctors in their daily practice. ML applications in cardiology are reflected by augmenting medical practice by contributing to early diagnosis, risk stratification, and personalized therapeutics. The examples of such applications in other domains include modeling disease progression of Alzheimer disease [3,4], Parkinson disease [5], multiple sclerosis [6], chronic kidney disease [7], chronic liver disease [8], and others.

Cardiovascular disorders in general are responsible for 30% of deaths worldwide. Among them specifically, hypertrophic cardiomyopathy (HCM) is a genetic cardiac disease that is a cause of sudden cardiac death (SCD), especially among young adults and athletes [9]. Cardiovascular diseases represent groups of diseases that can greatly benefit from preemptive prediction, prevention, and proactive management; thus, this opens an opportunity for methods of artificial intelligence (AI) [2]. Disease progression is especially hard to detect in slow-progressing diseases, such as HCM, which is present in about 1 of 500 young adults [10]. Although HCM has 4 identified stages [11], patients with HCM can experience a sudden cardiac arrest or the disease can slowly progress over several years. Currently, the state-of-the-art HCM Risk-SCD calculator method for risk stratification of patients diagnosed with HCM [12] is widely used in practice. Although this method predicts the risk of SCD, no methods, to the best of our knowledge, are available for modeling the patient's clinical status up to 10 years ahead. Detection of cardiovascular risk for 10 years ahead is important and has been recently modeled for atherosclerotic cardiovascular disease [13].

In this paper, we propose a novel ML-based tool for predicting disease progression for patients diagnosed with HCM in terms of adverse remodeling of the heart during a 10-year period. The method consists of 6 contemporaneous predictive regression models that independently predict future values of the following 6 clinical characteristics: left atrial diameter (LA\_d), left atrial volume (LA\_Vol), left ventricular ejection fraction (LVEF), New York Heart Association (NYHA) functional classification, left ventricular internal diameter at end diastole (LVIDd), and left ventricular internal diameter at end systole (LVIDs). Each prediction is supplemented with the explanation that is generated using the Shapley additive explanation (SHAP) method [14].

Comparison between current and future values of these 6 parameters, as well as the interpretation of the change, generated by explanation methods, can help cardiologists gain insight into the disease progression trend for a given patient.

### Machine Learning Methods in Medicine

ML techniques are being frequently applied in medicine to improve the prediction of disease progression, extraction of medical knowledge for outcome research, therapy planning and support, and overall patient management [15]. A wide variety of ML approaches can solve challenging problems in these tasks. For example, diseases such as Alzheimer disease, diabetes, and chronic obstructive pulmonary disease (COPD) progress slowly over the years. For modeling of COPD, a Markov model was proposed by Wang et al [16], who also included a database of virtual patients. Their method successfully modeled progression trajectories, showing that multiple progression trajectories are possible for some diseases.

In cardiology, there are several works addressing disease progression trends related to different cardiological diseases. With the increase in computational power, ML has become a tool to analyze nonlinear dependencies that are present either in relational data or in images. Juarez-Orozco et al [17] emphasized the advantages of ML, especially deep learning, in cardiac nuclear imaging, where ML can aid with ischemia diagnosis and event prognosis. Sardar et al [18] emphasized the advantages of AI in interventional cardiology, which is promising to bring a paradigm shift in the practice of medicine by improving real-time clinical decision making and standardizing robotic medical procedures. While focusing on the use of ML in electrocardiogram (ECG) analysis, Elul et al [19] also stated the crucial disadvantages of ML, which include a lack of explanation, relating the automated diagnosis with medical knowledge, and transparency of the system's limitations. In their work, the authors proceeded to flag individual predictions that are irrelevant or not useful. To summarize, the mentioned works characterize AI as a developing tool that, with the synergy between humans and machines, will help transform medical practice and clinical care.

Further, a hybrid approach for progression of Parkinson disease [5] was successfully used by combining a variety of ML methods from different families: clustering, dimensional reduction, and incremental support vector regression. Deep learning was used for predicting Alzheimer disease, on average, about 6 years in advance [20] and for modeling Alzheimer disease progression [4]. Conditional restricted Boltzmann machines were also used for prediction of disease progression

[3]. The authors simulated patient trajectories using 18 months of longitudinal data of around 1900 patients and showed that patient-level simulations are feasible using ML and appropriate data.

Several other ML approaches also model disease progression well in other medical domains, such as kidney disease progression [7]. In this work, 9 ML approaches were tested: linear regression (LR), elastic net regression, lasso regression, ridge regression, support vector machines (SVMs), random forests (RFs), k-nearest neighbors (KNNs), neural networks (NNs), and XGBoost. Similarly, ML models were applied to the problem of disease progression for hepatitis C virus [8] for the 5-year prediction problem using longitudinal data. The authors' conclusion was that the boosted survival tree-based models using longitudinal data perform better than cross-sectional or linear models. Last but not least, ML was also used for disease progression and secondary progression detection for multiple sclerosis [6]. Several ML models were evaluated for predictions of disease severity in 6-10 years, such as KNNs, decision trees, LR, and SVMs. SVMs performed best.

To summarize, the overview indicates that ML models can be successfully applied to problems of predicting disease progression, which is also the goal of this paper. In the next subsection, we overview how ML approaches are used in cardiology, specifically for HCM, which is the focus of this paper.

### Machine Learning for Modeling Hypertrophic Cardiomyopathy

Most ML contributions to cardiovascular medicine focus on risk stratification of patients. One of the biggest obstacles to using data for a broader variety of ML applications is that data are usually stored in diverse repositories, which are not readily usable for cardiovascular research, due to various data quality challenges [2]. Where the data are readily available, different ML algorithms have been successfully used, such as Wasserstein generative adversarial networks [21], convolutional NNs [22,23], deep NNs [24], and boosted decision trees [25]. Some authors have tested multiple models, such as RFs, artificial NNs, SVMs, and Bayesian networks [26], or a combination of J48, naive Bayes, KNNs, SVMs, RFs, bagging, and boosting [27]. Cuocolo et al [1] overviewed ML methods in cardiology, emphasizing their successful applications for building clinical predictive models, for analyzing ECG signals and image data. For the latter problems, the most successful methods were NNs, deep NNs, and convolutional networks. Advances in prediction accuracy have also been made by using deep NNs to make predictions based on fast, large-scale genome-wide association studies [28].

HCM is a severe disease for which 4 stages of progression have been identified in the medical literature [11]. Current state-of-the-art ML mostly uses only statistical models, such as multivariate regression analysis, which uses preselected predictor variables of known medical importance. Cardiac magnetic resonance (CMR) images [29,30] and echocardiographic diagnostics [31] are found to be a good source

of important attributes for HCM identification. Recently, researchers have started proposing ML-based risk stratification for patients diagnosed with comorbidities to separate patients into low- and high-risk categories or several categories on a scale [32]. The medical literature is mostly focused on finding risk factors that identify increased risk of SCD in patients with HCM [12,33]. A study [34] presenting the guidelines used in risk stratification for patients with HCM proposed potential SCD modifiers. Maron et al [35] performed a similar study on older populations and also summarized risk factors that could prevent SCD. The continuation of this research [36] aimed to develop an accurate strategy to assess the reliability of SCD prediction methods in prevention of SCD in patients diagnosed with HCM.

It is important to note that patients with HCM who experience cardiac arrest are not identified by typical risk markers used in the American College of Cardiology or the statistical mathematical risk model by the European Society of Cardiology [37]. Therefore, new risk factors have been and still need to be considered and developed to provide additional information to better assess HCM risk. In our work, we focus on modeling the future development of HCM by predicting the change in relevant cardiac parameters 10 years ahead.

### Aims and Contributions

Novelties and contributions of this paper include:

- A disease progression system that comprises models for prediction of 6 contemporaneous relevant clinical parameters that are relevant to HCM for 10 years ahead. The system includes the implementation of the explanation methodology that provides interpretability of predictive models.
- Analysis of predictive performance if training data are extended using semisupervised learning or with artificial patient data.
- Validation of predictive accuracy with medical experts by comparing ML and human accuracy and by analyzing sensibility of the computer-generated prediction explanations.

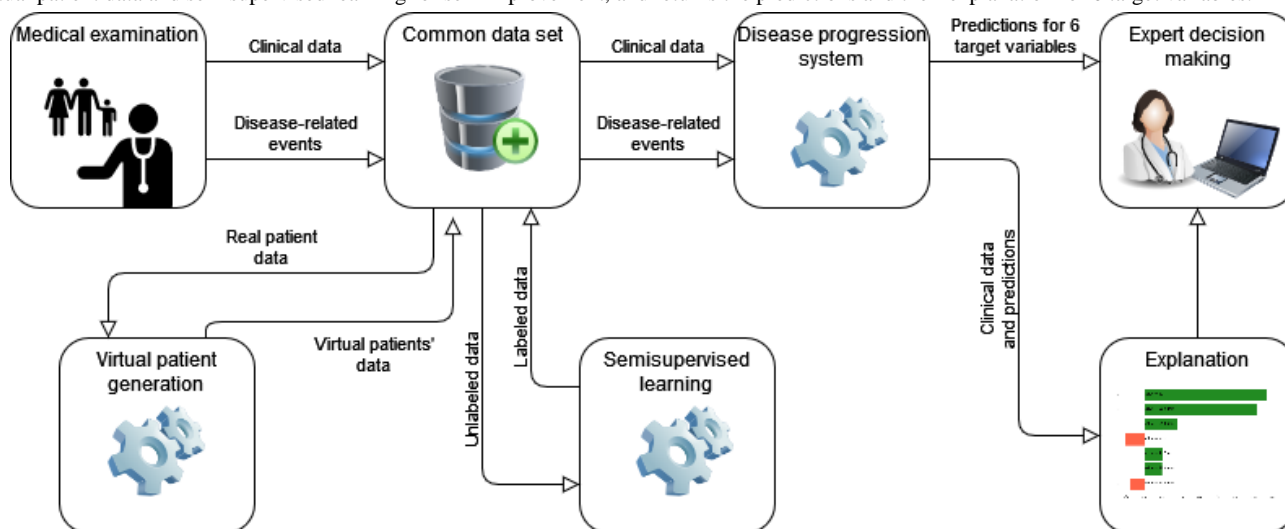
The aim of this paper is to develop a system capable of detecting slow progression of HCM based on longitudinal data.

## Methods

### Modeling Disease Progression

In this work, we modeled disease progression by predicting 6 relevant patient parameters 10 years in advance. These parameters are indicators of HCM and can be used to determine the stage of HCM according to the known guidelines [37]. Additionally, a preliminary analysis was performed to verify the prediction strength of the chosen parameters, validating our choice, as described in the Data Set section. The proposed disease progression system (Figure 1) takes as input patients' clinical data and data about their past disease-related events, such as dates of atrial fibrillation or syncope.

**Figure 1.** Overview of the proposed disease progression system. The system receives clinical data and disease-related events of a patient as input, uses virtual patient data and semisupervised learning for self-improvement, and returns the predictions and their explanation for 6 target variables.



The output of the system is a set of 6 contemporaneous target predictions for parameters:

- LA\_d
- LA\_Vol
- LVEF
- LVIDd
- LVIDs
- NYHA functional classification

In addition to predictions, the system also generates their explanations, revealing the factors with the largest impact on the increase or decrease in the 6 target variables throughout the 10-year period.

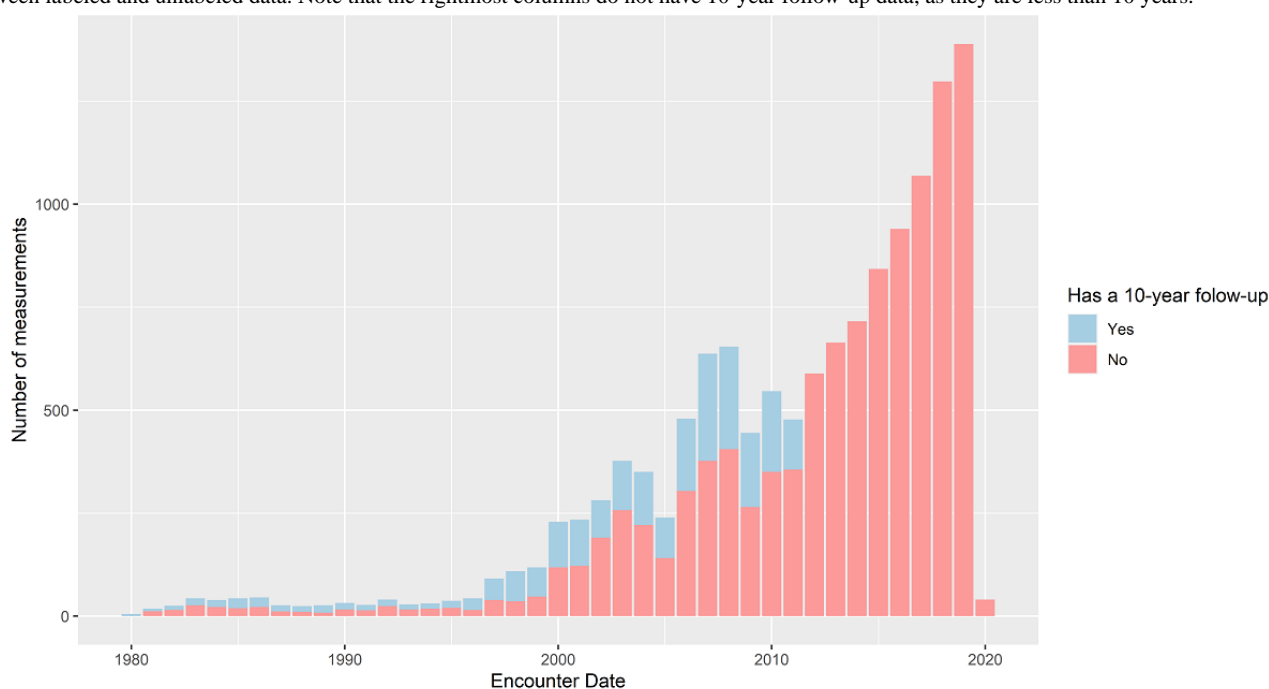
We trained the proposed disease progression system using supervised ML techniques. To further improve the results, we augmented the original data using unlabeled data (semisupervised learning) and virtual patients' data. We applied the semisupervised learning using patients without 10-year follow-ups and generated virtual patients' data using various techniques for artificial data generation. The semisupervised learning first predicted patients' targets using the trained models on labeled data, so they could be afterward included into the training data set. In the following subsections, we describe the data set, predictive modeling with supervised models, use of semisupervised learning and virtual patient data, and generation of prediction explanations.

### Data Set

The proposed approach was developed on a data set that was provided by the University of Florence as a result of its long-term clinical practice. The data set included patients who were enrolled over the past 40 years (Figure 2), and 1860 (80.24%) of 2318 patients had at least 1 available 10-year

follow-up. They were followed for an average duration of about 7 years and ranging up to 37 years. The data set contains longitudinal clinical data for 2318 patients diagnosed with HCM or patients that had a relative diagnosed with HCM (1457 [62.86%] male and 861 [37.14%] female patients). During the patients' visits, various clinical tests and relevant disease-related events were recorded. These data included general data (gender, age, height, weight, etc), genetic data (detected mutations), clinical tests (echocardiogram [echo], Holter monitoring, blood test, CMR, stress test), prescribed medications (type, start date, termination date), and disease-related events (eg, SCD, heart failure, transplant, abnormal Holter, pacemaker or implantable cardioverter defibrillator implantation). Echo was the leading diagnostic reference technique that was performed for the vast majority of patients and thus represents the main source of data. CMR was additionally used selectively due to its greater accuracy in measuring volumes. Although echo and CMR are treated separately and never computationally compared to each other in medical practice, we used CMR, where available, as an additional data modality to possibly improve prediction accuracy. In total, there were 6227 events recorded, of which 4902 (78.72%) events occurred in patients who were primarily diagnosed with HCM. The structure of the data set therefore allowed observing how patients' clinical characteristics change over time, which is essential for the desired modeling of HCM progression. The basic patient characteristics are shown in Table 1 for continuous parameters, Table 2 for binary parameters, and Table 3 for the remaining parameters. The characteristics were extracted from 10,318 measurements in total. Additionally, Table 4 shows the missing data numbers and percentages for the 6 selected target variables for their role as input or target variables.

**Figure 2.** Relationship between the amount of labeled and unlabeled data. The bars for Yes and No values are stacked, visually revealing the ratio between labeled and unlabeled data. Note that the rightmost columns do not have 10-year follow-up data, as they are less than 10 years.



**Table 1.** Basic characteristics of patients for basic continuous parameters (N=10,318).

Continuous parameter	Mean (SD)	Missing data, n (%)
Age (years)	52.1 (18.6)	4 (0.04)
Weight (kg)	73.4 (14.6)	2381 (23.08)
Height (cm)	169 (10.3)	2273 (22.03)
Body mass index (BMI)	25.6 (4.09)	2423 (23.48)
NYHA <sup>a</sup>	1.69 (0.73)	983 (9.53)

<sup>a</sup>NYHA: New York Heart Association.

**Table 2.** Basic characteristics of patients for basic binary parameters (N=10,318).

Binary parameter	1-value, n (%)	0-value, n (%)	Missing, n (%)
Alcohol	Yes, 103 (0.99)	No, 10,215 (99)	0
Drug	Yes, 18 (0.17)	No, 10,300 (99.83)	0
Smoking	Yes, 3437 (33.31)	No, 6881 (66.69)	0
Pregnancy	Yes, 443 (4.29)	No, 9875 (95.71)	2515 (24.37)
Gender	Male, 6400 (62.03)	Female, 3918 (37.97)	0

**Table 3.** Basic characteristics for groups of parameters (N=10,318)<sup>a</sup>.

Procedure	Parameters, n	Total missing values, n (%)
ECG <sup>b</sup>	9	45,839 (49.36)
Echo <sup>c</sup>	26	98,191 (36.60)
CMR <sup>d</sup>	10	81,174 (78.67)

<sup>a</sup>The table shows aggregated statistics for several parameters obtained from the same procedure. The percentage for each procedure is obtained as follows: [Total missing values/(Parameter × N)] × 100.

<sup>b</sup>ECG: electrocardiogram.

<sup>c</sup>Echo: echocardiogram.

<sup>d</sup>CMR: cardiovascular magnetic resonance.

**Table 4.** Absolute number and percentage of missing values of target variables as class and as input (N=10,318).

	LA_d <sup>a</sup> , n (%)	LVEF <sup>b</sup> , n (%)	NYHA <sup>c</sup> , n (%)	LVIDd <sup>d</sup> , n (%)	LVIDs <sup>e</sup> , n (%)	LA_Vol <sup>f</sup> , n (%)
Target	8569 (83.05)	8481 (82.19)	8313 (80.57)	8607 (83.42)	9336 (90.48)	8631 (83.65)
Input	2691 (26.08)	2399 (23.25)	983 (9.53)	2517 (24.39)	5329 (51.65)	3680 (35.67)

<sup>a</sup>LA\_d: left atrial diameter.

<sup>b</sup>LVEF: left ventricular ejection fraction.

<sup>c</sup>NYHA: New York Heart Association.

<sup>d</sup>LVIDd: left ventricular internal diameter at end diastole.

<sup>e</sup>LVIDs: left ventricular internal diameter at end systole.

<sup>f</sup>LA\_Vol: left atrial volume.

First, we transformed the available data set into a suitable form for predicting a 10-year change in relevant parameters using ML. Similarly, in other real-world data sets, most of the clinical tests were missing many patients or measurements were not taken for the whole span of 10 years (Figure 2). To address this issue, we preprocessed the data as follows:

- Formation of training examples: Since not all clinical tests can be conducted on the same day or in the same month, we defined a training example as a set of measurements within a time frame of 1 year. Such time frame corresponds to the annual regular visit period of patients and allows enough time for relevant changes in the observed parameters to become noticeable, as the disease slowly progresses. If the patient had a certain test performed multiple times within this time frame, multiple tests were treated as separate measurements. In case a certain type of test was not performed in the 1-year time frame, the corresponding variables were recorded as missing. Constructing training examples in this way yielded a data set with 13,386 examples, with 3.9 (SD 4.8) examples per patient.
- Imputation of missing data: The missing values in the data set, either because of nonperformed tests or because of erroneous input of data, were imputed by copying the closest past values (sensible because the progression of HCM is slow; used on numerical and categorical attributes), imputing values of a healthy patient (sampled from the normal distribution; used for numerical attributes), or

imputing mean values where healthy values were unknown (used on numerical and categorical attributes). Since measurements were not taken at equidistant time intervals, we used linear interpolation for computing equidistant measurement approximations.

We used the formed training examples as input to supervised learning algorithms. Prior to modeling, we evaluated the quality of attributes, which is important for decreasing learning complexity, avoiding overfitting, and, therefore, improving the simplicity and performance of ML methods. To facilitate learning with NNs, we also scaled the values to the interval [0,1] and encoded nominal values using the one-hot encoding method.

We used RReliefF [38], adaptation of the ReliefF feature selection algorithm, for regression problems. RReliefF calculates how well a feature's values distinguish between distant labels of instances that are close to each other and considers feature interactions. We selected 21 (18.7%) of 112 attributes based on the average rank across all 6 target variables for further supervised learning. Feature scores for 21 selected features are shown in Table 5, along with their average ranks across 6 trained predictive models. After removing highly correlated features (eg, the weight feature that correlates to the body surface area [BSA] and height), the final set of attributes contained all target variables (regardless of their rank) and the best-performing attributes, based on average rank.

**Table 5.** Selected attributes using RReliefF.<sup>a</sup>

Variable <sup>b</sup>	LA_d <sup>c</sup> score	LVEF <sup>d</sup> score	NYHA <sup>e</sup> score	LVIDd <sup>f</sup> score	LVIDs <sup>g</sup> score	LA_Vol <sup>h</sup> score	Average rank
<b>Anthropometric parameters</b>							
<i>Age</i>	0.198	0.194	0.166	0.142	0.166	0.158	1.000
<i>Gender</i>	0.051	0.037	0.043	0.055	0.058	0.022	12.500
<i>Height</i>	0.057	0.064	0.045	0.075	0.051	0.029	9.167
<i>BSA</i> <sup>i</sup>	0.075	0.073	0.053	0.095	0.085	0.045	4.167
<b>Risk factors</b>							
<i>Smoking</i>	0.063	0.046	0.052	0.032	0.069	0.082	7.500
<i>Presence of hypercholesterolemia</i>	0.072	0.042	0.052	0.039	0.044	0.056	9.667
History of syncope	0.026	0.036	0.029	0.022	0.029	0.048	20.000
<i>Family history of HCM</i> <sup>j</sup>	0.056	0.060	0.061	0.047	0.052	0.066	5.833
Family history of SCD <sup>k</sup>	0.027	0.051	0.032	0.031	0.051	0.049	14.667
<b>Clinical, ECG<sup>l</sup>, and echo<sup>m</sup> parameters</b>							
NYHA	0.011	0.017	0.069	0.007	0.027	0.022	33.000
Presence of atrial fibrillation	0.055	0.036	0.048	0.018	0.026	0.068	16.333
QRS duration	0.035	0.046	0.029	0.039	0.026	0.039	17.167
<i>Interventricular septum ( IVS)</i>	0.043	0.052	0.049	0.041	0.057	0.052	8.167
LA_d	0.078	0.037	0.036	0.018	0.031	0.070	15.000
LA_Vol	0.055	0.029	0.026	0.012	0.025	0.059	24.000
LVIDs	0.017	0.022	0.027	0.029	0.043	0.031	25.167
LVIDd	0.021	0.017	0.017	0.036	0.044	0.026	27.667
LVEF	0.018	0.051	0.019	0.014	0.050	0.013	27.833
<b>Genetics</b>							
<i>Mutation MYBPC3</i>	0.045	0.041	0.039	0.051	0.052	0.059	9.667
<i>Mutation MYH7</i>	0.037	0.044	0.034	0.040	0.066	0.023	14.667
Negative genetics	0.036	0.037	0.027	0.043	0.030	0.031	18.667

<sup>a</sup>The table shows RReliefF feature scores and the average ranks for each target variable.

<sup>b</sup>Names of the 10 highest-ranked variables are italicized.

<sup>c</sup>LA\_d: left atrial diameter.

<sup>d</sup>LVEF: left ventricular ejection fraction.

<sup>e</sup>NYHA: New York Heart Association.

<sup>f</sup>LVIDd: left ventricular internal diameter at end diastole.

<sup>g</sup>LVIDs: left ventricular internal diameter at end systole.

<sup>h</sup>LA\_Vol: left atrial volume.

<sup>i</sup>BSA: body surface area.

<sup>j</sup>HCM: hypertrophic cardiomyopathy.

<sup>k</sup>SCD: sudden cardiac death.

<sup>l</sup>ECG: electrocardiogram.

<sup>m</sup>Echo: echocardiogram.

## Predictive Modeling With Supervised and Semisupervised Machine Learning

To model the relationship between input patient data and target variables, we applied the following supervised learning algorithms:

- RFs [39,40] are an ensemble prediction model that construct multiple randomized decision trees. The implementations of an RF classifier in the R statistical package (library *ranger*) and the Python Scikit-Learn package [41] were used. Each forest used between 500 and 1500 trees, and the Gini index was used as the attribute-splitting rule.
- Gradient boosting (XGBoost) [42] is an ensemble of weak decision tree predictors, implemented in the open source software library XGBoost.
- LR is a traditional method of finding a linear dependence between attributes and the selected target variable.
- NNs mimic the architecture and working of brain neurons. We used 1 input and 1 output layer and 1 or several hidden layers. In the optimization process, we optimized several learning parameters, such as the learning rate, number of hidden layers, sizes of layers, regularization, sample weights, class weights, dropout, and batch normalization.

The best hyperparameters of these algorithms were tuned using Bayesian optimization and random search implemented in *keras-tuner* [43].

## Semisupervised Learning and Virtual Patients

Semisupervised learning is increasingly used in medicine, especially for medical image segmentation [44-46]. This approach allows labeling a large amount of unlabeled data using only a small portion of labeled data. The majority (ie, 83.9% averaged over 6 target variables) of patients' data did not have records for the follow-up after 10 years. These unlabeled data were used as examples for semisupervised learning, producing a teacher model. The unlabeled examples were labeled with the supervised learning predictive model (see the Predictive Modeling With Supervised and Semisupervised Machine Learning section) and added to the training set. After that, a new model (also called a student model) was trained and kept if it achieved better performance on the test set than the teacher model.

To further improve the results of semisupervised learning, we used artificially generated data (ie, virtual patients). Virtual data generation can sometimes replace experiments in biomedical experiments on animals [47]. Specifically in cardiovascular modeling, patient-specific virtual patient modeling has recently made major progress in improving diagnoses [48]. We evaluated the performance and appropriateness of several virtual patient data generators for this task, such as the generator based on the multivariate normal and log-normal distribution (MVND and log-MVND) [49], and nonparametric methods using supervised tree ensembles, unsupervised tree ensembles, radial basis function-based NNs [50], and Bayesian networks [51]. As the final data generator, we chose the unsupervised tree ensembles, which exhibited the highest level of agreement between the real and the virtual distributions, computed with the Kolmogorov-Smirnoff goodness-of-fit statistical test [52]. We

generated 10,000 virtual patient examples, with 20 most important features, listed in Data Set section.

## Explanation of the Predictive Model

Supervised ML models often exhibit a black-box nature, meaning that they can model data but not provide an explanation for the contained knowledge as well as the reasoning used in predictions. This means that the models lack transparency and interpretability. To address this, explanation methods provide justification for each prediction and assess features with the highest impact [53]. This is important in risk-sensitive ML application areas, such as medicine, where the predictions of ML models need to be understood as they may represent a basis for further medical interventions.

In our work, we applied the SHAP method [14], which is a model-agnostic method, generating an explanation for different ML models in a unified form. The method uses theoretically sound concepts of Shapley values from cooperative game theory for computing contributions of each individual attribute value and of each attribute overall. The generated explanations visualize the most relevant attributes that contribute to higher or lower prediction values. The explanations can be computed either for a single patient's predictions or summarized over all patients to discover more general relationships between attributes and the model's predictions.

## Results

### Models' Comparison

To evaluate and compare the performance of the 6 predictive models, we used stratified 10-fold cross-validation. For each of the 6 predictive problems, 4 different regression models were evaluated (LR, RF, gradient-boosted [GB] trees, and NN). The following parameters were varied in tests:

- Application of semisupervised learning (denoted with S)
- Addition of virtual patients' data into the learning data set (denoted with VP)
- Use of all 112 features (denoted with All) or only a subset of the 21 best features (denoted with Subset)
- Interpolation of data points so that measurements were equidistant (denoted with I)

In all, 28 different combinations of the parameters were used in experiments. Some combinations were omitted due to limitations (eg, VP generators cannot generate data for all 112 attributes, so VP was evaluated only with the subset of attributes) or excessive time complexity (eg, the use of virtual patients with NNs).

### Performance of Predictive Models

To compare the accuracy of the obtained models, we computed the following 4 metrics: mean absolute error (MAE), root-mean-square error (RMSE), and 2 variations of the relative root-mean-square error (RRMSE<sub>mean</sub> and RRMSE<sub>const</sub>). The MAE measures the average absolute difference between predicted and true values over all examples in the test set. The RMSE addresses the issue that the squared values of the MSE are hard to interpret. The RRMSE measures the relative ratio between the obtained model and the baseline model. We



computed 2 variations of the RRMSE with 2 different baseline models: mean predictor and constant predictor. With the  $RRMSE_{mean}$ , we compared the performance of the obtained model to the model that returned the mean of the target variable over all patients (mean predictor), while with the  $RRMSE_{const}$ , we compared the obtained model to the model that assumed that the value of the target variable would remain constant/unchanged over the 10-year period (constant predictor).

We summarized (Table 6) the performance of the best-performing predictive models (RF, LR, GB, NN) and parameters (S, VP, All/Subset) for each target variable. We could see that the top-performing regression models were the RF and the GB tree for all target variables. We achieved the best results by applying semisupervised learning (S) for all target variables and using virtual patients (VP) for 5 of 6 target variables. For all targets, the best results were obtained by learning from a subset of the 21 most important features. The values of both RRMSE metrics revealed that the model performs

better than the baseline models (their values are less than 1.0), with the model for the LA\_d target achieving the lowest predictive error.

To further evaluate the contribution of different data augmentation strategies, we compared the results on different patient sets: original (All features), subset of best features (Subset), virtual patients (VP), semisupervised learning (S), and the combination of the latter 2 (S + VP). The obtained results, shown for the best-performing RF model, are given in Figure 3, which compares the  $R^2$  metrics for each individual target parameter. The additional detailed results for the other models are given in Multimedia Appendix 1. The obtained results reveal the benefits of reducing the feature space, as well as applying the used data augmentation methods.

In the following subsection, we apply the explanation methodology that helps interpret the computed predictions and their contributing feature values.

**Table 6.** Comparison of the best-performing models for each target variable.

Target	Model and parameter	MAE <sup>a</sup>	RMSE <sup>b</sup>	$RRMSE_{mean}^c$	$RRMSE_{const}$
LA_d <sup>d</sup>	RF <sup>e</sup> : S <sup>f</sup> +VP <sup>g</sup> +Subset	3.4	4.73	0.54	0.46
LA_Vol <sup>h</sup>	RF: S+VP+Subset	18.4	26.73	0.56	0.47
LVEF <sup>i</sup>	GB <sup>j</sup> : S+Subset	4.92	6.73	0.67	0.61
LVIDd <sup>k</sup>	RF: S+VP+Subset	3.53	5.26	0.68	0.64
LVIDs <sup>l</sup>	RF: S+VP+Subset	3.42	4.81	0.66	0.56
NYHA <sup>m</sup>	RF: S+VP+Subset	0.39	0.5	0.67	0.66

<sup>a</sup>MAE: mean absolute error.

<sup>b</sup>RMSE: root-mean-square error.

<sup>c</sup>RRMSE: relative root-mean-square error.

<sup>d</sup>LA\_d: left atrial diameter.

<sup>e</sup>RF: random forest.

<sup>f</sup>S: application of semisupervised learning.

<sup>g</sup>VP: addition of virtual patients' data into the learning data set.

<sup>h</sup>LA\_Vol: left atrial volume.

<sup>i</sup>LVEF: left ventricular ejection fraction.

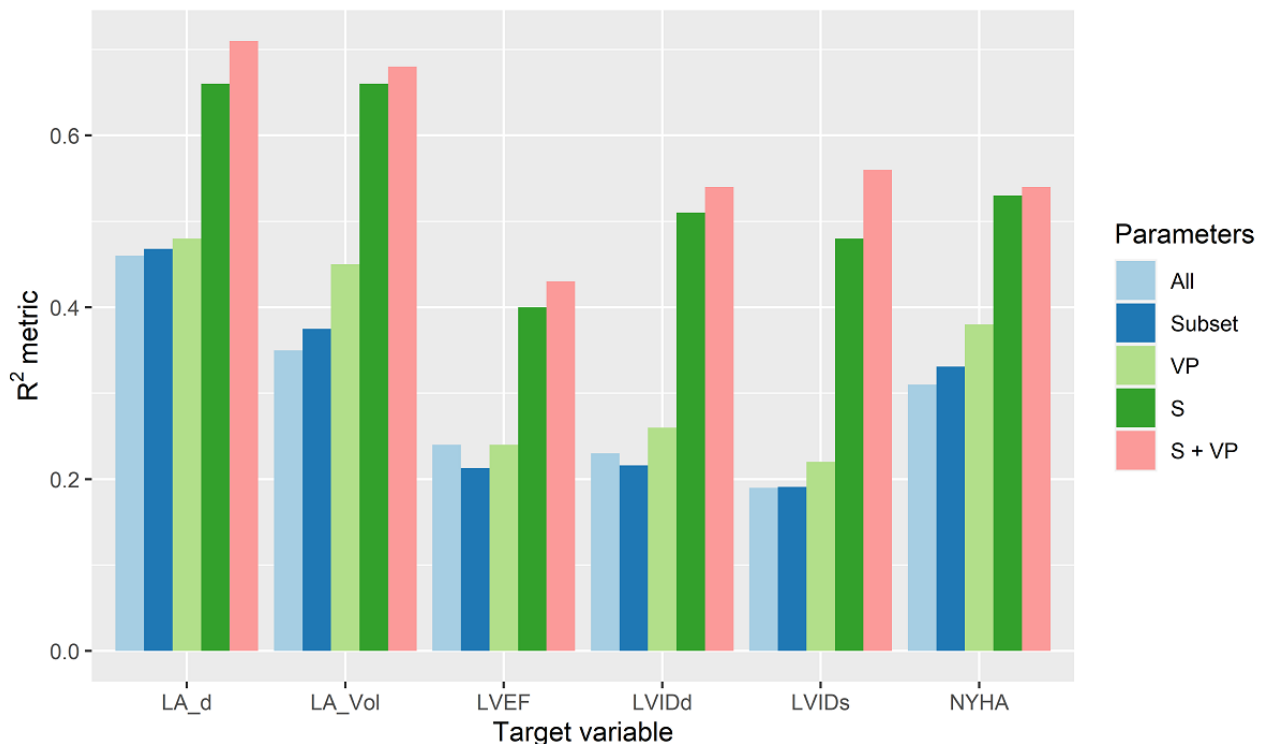
<sup>j</sup>GB: gradient boosted.

<sup>k</sup>LVIDd: left ventricular internal diameter at end diastole.

<sup>l</sup>LVIDs: left ventricular internal diameter at end systole.

<sup>m</sup>NYHA: New York Heart Association.

**Figure 3.** Plotted results for the  $R^2$  statistic for each target variable using different sets (input parameters). Note that VP, S, and S + VP are used on feature subsets. LA\_d: left atrial diameter; LA\_Vol: left atrial volume; LVEF: left ventricular ejection fraction; LVIDd: left ventricular internal diameter at end diastole; LVIDs: left ventricular internal diameter at end systole; NYHA: New York Heart Association; S: application of semisupervised learning; VP: addition of virtual patients' data into the learning data set.



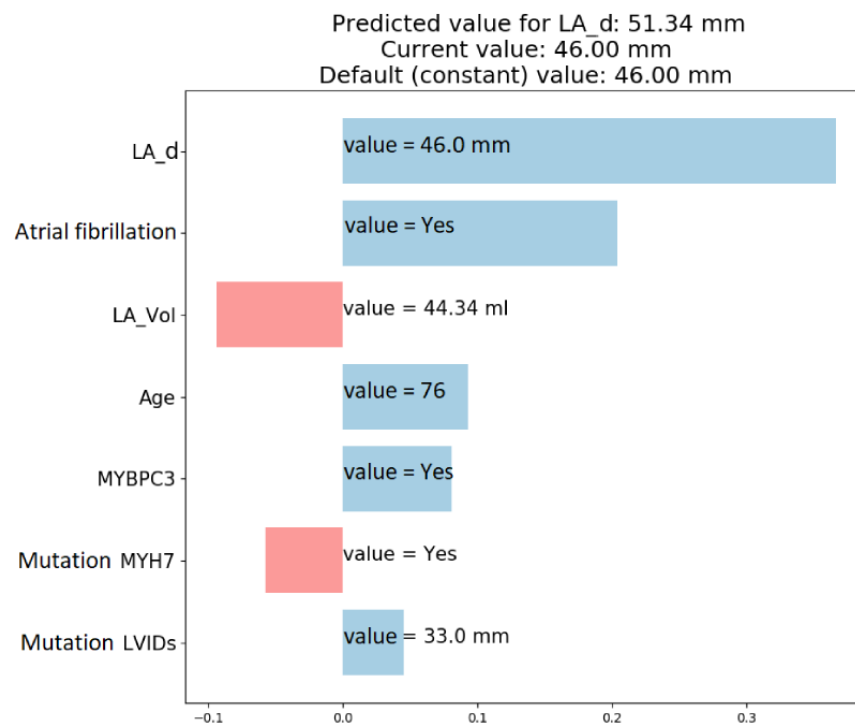
### Explanation of Predictions

To augment the output of prediction models, we applied the SHAP method [14] for computing explanations of individual predictions. The explanation of a single prediction consists of relevant textual, graphical, and numerical data that allows understanding of the relationships between the features of the patient and the model's prediction. It also consists of a list of the most relevant features that influence the prediction, along with their contribution values that define whether the feature value either supports the predicted value or opposes it. The direction of the impact (ie, sign of the contribution value) is denoted using different colors.

An example of the explanation generated for the prediction for the target LA\_d (Figure 4) is presented here. Features' contributions are sorted in descending order, and the graph contains only the features for which the sum of their

contributions reflects 95% of the difference between the initial parameter value and the predicted value after 10 years. The green and red bars thus denote positive and negative contributions of the impact for individual feature values, respectively, showing the factors contributing to the increase or decrease in the LA\_d value. We can see that the LA\_d, atrial fibrillation, age, mutation MYBPC3, and LVIDs features contributed to the increase in the predicted value for LA\_d over time, while the LA\_Vol and mutation MYH7 features contributed to the decrease in the predicted value for LA\_d. Because the overall increasing impact was more prominent, the final predicted value (51.34) was higher than the baseline prediction, which is also the current patients' value of LA\_d (46.00). Larger magnitudes of the features' contributions correspond to larger changes in the prediction value. For example, LA\_d contributed the most (approximately 30%) to the increase in the predicted value.

**Figure 4.** Example of an explanation of the prediction for the target variable LA\_d. LA\_d: left atrial diameter; LA\_Vol: left atrial volume; LVIDs: left ventricular internal diameter at end systole.



### Validation With Medical Experts

Besides evaluation of prediction models with statistical measures conducted in 2 previous sections, we engaged medical experts to provide further interpretation and validation of the results. First, we compared the accuracy of predictive models with the accuracy of human experts, which was obtained by using a survey (Multimedia Appendix 2). Second, we checked whether prediction explanations were sensible and consistent with the experts' medical knowledge about HCM.

We prepared a questionnaire for medical experts and distributed it to several medical universities and cardiology clinics. The questionnaire included data about complete medical cases (measurements, events, and medication data) for 10 patients, and the experts were asked to study them and complete the following 2 tasks:

- Predict the magnitude of the 10-year change in the 6 studied clinical parameters (LA\_d, LA\_Vol, LVEF, LVIDd, LVIDs, and NYHA) and mark it on a discrete scale from  $-3$  to  $3$ , where  $-3$  and  $3$  represented the biggest-possible decrease and increase, respectively. Possible magnitudes of change were represented using discrete intervals, as the prediction of an exact value is a difficult task that does not take place in medical practice.
- Evaluate whether the statements generated from the explanation (eg, "The current value of parameter LA\_d will cause a decrease in LA\_d") are true or false. For each patient, 6 such statements were generated, covering the features with the highest contribution. More specifically, the questionnaire included evaluation questions for 6

parameters that contribute to a change in LA\_d, 4 for LA\_Vol, 5 for LVEF, 6 for LVIDd, 7 for LVIDs, and 4 for NYHA.

The questionnaire was fully completed by 13 experts with 16 (SD 8) years of experience. In the following subsections, we present the analysis of the answers.

### Validation of Prediction Accuracy

To compare the prediction accuracy between the experts and the ML model, we first discretized the model's predictions into discrete intervals so that they could be compared to the discrete intervals, predicted by the experts. We performed the discretization using bins of width  $0.25\sigma$ , where  $\sigma$  is the SD of the variable. Further, we calculated the following prediction errors:

- Mean prediction error of the discretized model prediction (denoted with MD)
- Mean prediction error made by individual medical experts (denoted with E)
- Mean prediction error of the consortium prediction (ie, the average prediction of all doctors, denoted with C)

We could see that the mean prediction error of the discretized model MD (Table 7) was the lowest for all target variables except for LA\_d. The mean errors of consortium predictions C were lower than the predictions of individual experts for all parameters, which indicates that the mutual consolidation of different doctors' opinions reduced the error of their joint predictions. The consortium prediction error also turned out to be the lowest for the parameter LA\_d and thus better than the error of the ML model.

**Table 7.** Mean absolute error (MAE) of the discretized model predictions (MD), individual experts (E), and the entire consortium (C).

Target/prediction	Model (MD), MAE (SD)	Expert (E), MAE (SD)	Consortium (C), MAE (SD)
NYHA <sup>a</sup>	<i>0.30 (0.48)</i> <sup>b</sup>	0.84 (0.69)	0.56 (0.34)
LA_d <sup>c</sup>	1.70 (0.82)	1.69 (0.97)	<i>1.66 (0.70)</i> <sup>b</sup>
LA_Vol <sup>d</sup>	<i>1.00 (0.82)</i> <sup>b</sup>	1.25 (0.98)	1.13 (0.63)
LVIDd <sup>e</sup>	<i>0.80 (0.63)</i> <sup>b</sup>	1.09 (0.91)	1.00 (0.77)
LVIDs <sup>f</sup>	<i>0.50 (0.71)</i> <sup>b</sup>	1.02 (0.86)	0.88 (0.68)
LVEF <sup>g</sup>	<i>0.90 (0.88)</i> <sup>b</sup>	1.32 (0.90)	1.28 (0.79)

<sup>a</sup>NYHA: New York Heart Association.

<sup>b</sup>The lowest achieved errors are italicized.

<sup>c</sup>LA\_d: left atrial diameter.

<sup>d</sup>LA\_Vol: left atrial volume.

<sup>e</sup>LVIDd: left ventricular internal diameter at end diastole.

<sup>f</sup>LVIDs: left ventricular internal diameter at end systole.

<sup>g</sup>LVEF: left ventricular ejection fraction.

### Validation of the Model Explanation

To validate the generated model explanations, we analyzed the agreement of experts with statements generated about the features' influence in 2 steps. First, we calculated the agreement ratio for individual features that were included in the questionnaire, grouped by each of the 6 target variables. Second, we calculated the overall agreement of experts with the explanation for each of the 6 target parameters, based on the agreement data about all features that contributed to their prediction.

The results (Table 8) of the analysis provided the ratio of agreement between different parameters for each target variable, as well as their overall agreement. The highest agreement ratio was achieved for target attributes NYHA (1.00), LA\_Vol (0.75), and LVIDd (0.67). The last column (Average agreement) summarizes the results across all used features. The results, in decreasing order, of the last column show that the majority of the experts agreed, especially with the explanations for the targets NYHA (average agreement of 0.73) and LVIDd (average agreement of 0.52). By comparing Tables 7 and 8, we consistently see that the experts least agreed with explanations for the target LA\_d, for which the predictive model achieved a

larger error than individual experts or the entire consortium. In cases where the predictive model achieved better predictive accuracy than the experts (Table 7) and the agreement of the experts with the explanation was lower (Table 8), for example, for LVEF, LA\_Vol, and LVIDs, there are 3 possible explanations:

- The generated explanation might, indeed, provide incorrect information.
- The generated explanation might explain novel relationships between features and target parameters that have not been observed or documented so far.
- It was hard for the experts to evaluate the claims in the questionnaire about the influence of particular features, as these tasks deviate from the established medical practice and require the experts to rely on their subjective experience.

For establishing the reasons for imperfect agreement between the explanation and the experts, further investigation is therefore required. We can conclude that the results provide some evidence that the generated prediction explanation might provide a complementary view at the prediction of HCM-related parameters. Such explanations might represent a tool that the experts could consult while making their decisions.

**Table 8.** Agreement ratios between experts and prediction explanations for parameters that contribute to predicting each target variable. The last two columns provide summary statistics.

Target variable and parameters	Expert agreement	Summary Ratio of agreed features from at least 50% of experts, n	Average agreement, n
<b>NYHA<sup>a</sup></b>			
<i>LA_d<sup>b</sup></i>	0.77 <sup>c</sup>		
<i>Age</i>	0.77 <sup>c</sup>	1.00 (4/4)	0.73
<i>LA_Vol<sup>d</sup></i>	0.62 <sup>c</sup>		
<i>Atrial fibrillation</i>	0.77 <sup>c</sup>		
<b>LVIDd<sup>e</sup></b>			
<i>BSA<sup>f</sup></i>	0.15		
<i>Gender</i>	0.85 <sup>c</sup>		
<i>LVIDd</i>	0.65 <sup>c</sup>	0.67 (4/6)	0.52
<i>QRS duration</i>	0.69 <sup>c</sup>		
<i>LVEF<sup>g</sup></i>	0.23		
<i>Mutation MYH7</i>	0.54 <sup>c</sup>		
<b>LVEF</b>			
<i>QRS duration</i>	0.38		
<i>Presence of hypercholesterolemia</i>	0.54 <sup>c</sup>		
<i>Syncope</i>	0.46	0.40 (2/5)	0.49
<i>Gene_Testing_Performed</i>	0.69 <sup>c</sup>		
<i>NYHA</i>	0.38		
<b>LA_Vol</b>			
<i>LA_Vol</i>	0.69 <sup>c</sup>		
<i>BSA</i>	0.54 <sup>c</sup>	0.75 (3/4)	0.48
<i>Age</i>	0.15		
<i>Atrial fibrillation</i>	0.54 <sup>c</sup>		
<b>LVIDs</b>			
<i>LA_d</i>	0.38		
<i>LVIDd</i>	0.38		
<i>LA_Vol</i>	0.62 <sup>c</sup>		
<i>BSA</i>	0.85 <sup>c</sup>	0.43 (3/7)	0.47
<i>Mutation MYBPC3</i>	0.62 <sup>c</sup>		
<i>Interventricular septum (IVS)</i>	0.38		
<i>Family history of HCM<sup>h</sup></i>	0.08		
<b>LA_d</b>			
<i>LA_d</i>	0.85 <sup>c</sup>		
<i>Atrial fibrillation</i>	0.15		
<i>BSA</i>	0.08	0.17 (1/6)	0.36
<i>IVS</i>	0.38		

Target variable and parameters	Expert agreement	Summary
		Ratio of agreed features from at least 50% of experts, n
		Average agreement, n
Age	0.31	
LVEF	0.38	

<sup>a</sup>NYHA: New York Heart Association.

<sup>b</sup>LA\_d: left atrial diameter.

<sup>c</sup>Names of parameters with agreement higher than 50% are italicized.

<sup>d</sup>LA\_Vol: left atrial volume.

<sup>e</sup>LVIDd: left ventricular internal diameter at end diastole.

<sup>f</sup>BSA: body surface area.

<sup>g</sup>LVEF: left ventricular ejection fraction.

<sup>h</sup>HCM: hypertrophic cardiomyopathy.

## Discussion

### Principal Results

We presented a disease progression system for patients diagnosed with HCM that is based on predicting 6 target parameters (LA\_d, LA\_Vol, LVIDd, LVIDs, LVEF, and NYHA) for 10 years ahead using supervised ML models. The experiments revealed good ML performance for all targets, with the achieved predictive error lower than the error of the default predictors. The experiments also revealed that semisupervised learning and the artificial data from virtual patients helped achieve even higher predictive accuracy for all 6 targets. Finally, we validated our approach with human experts using a structured questionnaire and determined the models' favorable performance compared to performance of experts for 5 of 6 targets.

### Limitations

The design of the study carried several limitations, stemming from the fact that this work was based on real-world data that are expensive to obtain and are subject to noise. The first limitation of this study is that it was based only on a single medical center data set. To further validate this study, it would be beneficial to independently evaluate the models with data sets from other centers or extend the existing data set with more data. Additionally, the benefit for including more data could also be in diminishing a potential bias of our data set, which could potentially include a population distribution that is different from other medical centers and thus different ranges of recorded parameters, which we did, in fact, observe in some cases. Additionally, in the perfect but rather unrealistic scenario due to its cost, both data modalities (echo and CMR) would be available for all patients, which would allow us to use the CMR data as an additional data source for all patients. Due to the unavailability of such data at the time of the study or data that were structured differently, we leave this for our further work.

Further, to prepare the data to be used for ML and obtain stable predictions, we used several preprocessing and data augmentation steps. Since we are dealing with real medical data, this opens questions of how different data transformations influence our predictions. Hence, a sensitivity study of the

results would be required, as well as determining how the patient's record time frame and predicted risk time frame influence the achieved accuracies. An additional limitation of the performed validation was that the ML results were compared to the inputs of medical experts in the structured survey instead of their free diagnoses and evaluations. Although this was required to unify the structure of human answers to enable statistical comparisons, the form of survey might introduce its own bias.

The described limitations, along with our further research questions and ideas, open several ideas for future study directions. First, we will evaluate the proposed system on an independent cardiological data set (eg, the Sarcomeric Human Cardiomyopathy Registry [SHaRe]) [54]. Second, as our current approach provides future predictions for 6 independent parameters, the outputs will be further combined into a single risk prediction of high/low risk, which can further improve HCM health management initiative [32]. To achieve this, a combination of models' output analysis and domain experts' input would be required. Finally, further ways for improvement of predictive accuracy will be tested (additional predictive models and feature selection techniques, including deep learning), as well as determining the reasons for the experts' disagreement with some of the explanation components.

### Conclusion

Although ML can have limitations in medicine [2], in this work, we showed the importance of using computer models in cardiology by predicting disease progression of HCM patients 10 years ahead, which could be used to prevent SCD. Additionally, the results confirmed findings in Chen et al [44], Gu et al [45], and Bai et al [46] that additional artificial data and semisupervised learning can provide additional low-cost and low-risk data using already available medical knowledge, increasing the predictive performance. Simple explanations of predictions contribute to the trust of provided predictions and ease the decision of experts. We hope that our work will further contribute to the goal of developing constructive strategies to prevent SCD in patients with HCM, as motivated by Maron et al [36].

## Acknowledgments

This project received funding from the European Union's Horizon 2020 research and innovation program (grant agreement no. 777204; [www.silicofcm.eu](http://www.silicofcm.eu)). This paper reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information it contains.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Bar graphs of parameter influence for each model used.

[\[PNG File , 257 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

A sample of the questionnaire for the first patient.

[\[PDF File \(Adobe PDF File\), 597 KB-Multimedia Appendix 2\]](#)

## References

1. Cuocolo R, Perillo T, De Rosa E, Ugga L, Petretta M. Current applications of big data and machine learning in cardiology. *J Geriatr Cardiol* 2019 Aug;16(8):601-607 [[FREE Full text](#)] [doi: [10.11909/j.issn.1671-5411.2019.08.002](https://doi.org/10.11909/j.issn.1671-5411.2019.08.002)] [Medline: [3155327](https://pubmed.ncbi.nlm.nih.gov/3155327/)]
2. Shameer K, Johnson K, Glicksberg B, Dudley J, Sengupta P. Machine learning in cardiovascular medicine: are we there yet? *Heart* 2018 Jul;104(14):1156-1164. [doi: [10.1136/heartjnl-2017-311198](https://doi.org/10.1136/heartjnl-2017-311198)] [Medline: [29352006](https://pubmed.ncbi.nlm.nih.gov/29352006/)]
3. Fisher CK, Smith AM, Walsh JR, Coalition Against Major Diseases. Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci Rep* 2019 Sep 20;9(1):13622 [[FREE Full text](#)] [doi: [10.1038/s41598-019-49656-2](https://doi.org/10.1038/s41598-019-49656-2)] [Medline: [31541187](https://pubmed.ncbi.nlm.nih.gov/31541187/)]
4. Lee G, Nho K, Kang B, Sohn K, Kim D. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci Rep* 2019 Feb 13;9(1):1952 [[FREE Full text](#)] [doi: [10.1038/s41598-018-37769-z](https://doi.org/10.1038/s41598-018-37769-z)] [Medline: [30760848](https://pubmed.ncbi.nlm.nih.gov/30760848/)]
5. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L, Farahmand M. A hybrid intelligent system for the prediction of Parkinson's disease progression using machine learning techniques. *Biocybern Biomed Eng* 2018;38(1):1-15. [doi: [10.1016/j.bbe.2017.09.002](https://doi.org/10.1016/j.bbe.2017.09.002)]
6. Pinto MF, Oliveira H, Batista S, Cruz L, Pinto M, Correia I, et al. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Sci Rep* 2020 Dec 03;10(1):21038 [[FREE Full text](#)] [doi: [10.1038/s41598-020-78212-6](https://doi.org/10.1038/s41598-020-78212-6)] [Medline: [33273676](https://pubmed.ncbi.nlm.nih.gov/33273676/)]
7. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *J Transl Med* 2019 Apr 11;17(1):119-113 [[FREE Full text](#)] [doi: [10.1186/s12967-019-1860-0](https://doi.org/10.1186/s12967-019-1860-0)] [Medline: [30971285](https://pubmed.ncbi.nlm.nih.gov/30971285/)]
8. Konerman MA, Beste LA, Van T, Liu B, Zhang X, Zhu J, et al. Machine learning models to predict disease progression among veterans with hepatitis C virus. *PLoS One* 2019 Jan 4;14(1):e0208141 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0208141](https://doi.org/10.1371/journal.pone.0208141)] [Medline: [30608929](https://pubmed.ncbi.nlm.nih.gov/30608929/)]
9. Lyon A, Mincholé A, Bueno-Orovio A, Rodriguez B. Improving the clinical understanding of hypertrophic cardiomyopathy by combining patient data, machine learning and computer simulations: a case study. *Morphologie* 2019 Dec;103(343):169-179 [[FREE Full text](#)] [doi: [10.1016/j.morpho.2019.09.001](https://doi.org/10.1016/j.morpho.2019.09.001)] [Medline: [31570308](https://pubmed.ncbi.nlm.nih.gov/31570308/)]
10. Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary artery risk development in (young) adults. *Circulation* 1995 Aug 15;92(4):785-789. [doi: [10.1161/01.cir.92.4.785](https://doi.org/10.1161/01.cir.92.4.785)] [Medline: [7641357](https://pubmed.ncbi.nlm.nih.gov/7641357/)]
11. Olivotto I, Cecchi F, Poggesi C, Yacoub MH. Patterns of disease progression in hypertrophic cardiomyopathy: an individualized approach to clinical staging. *Circ Heart Fail* 2012 Jul 01;5(4):535-546. [doi: [10.1161/CIRCHEARTFAILURE.112.967026](https://doi.org/10.1161/CIRCHEARTFAILURE.112.967026)] [Medline: [22811549](https://pubmed.ncbi.nlm.nih.gov/22811549/)]
12. O'Mahony C, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi C, Hypertrophic Cardiomyopathy Outcomes Investigators. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J* 2014 Aug 07;35(30):2010-2020. [doi: [10.1093/eurheartj/eh439](https://doi.org/10.1093/eurheartj/eh439)] [Medline: [24126876](https://pubmed.ncbi.nlm.nih.gov/24126876/)]
13. Andy AU, Guntuku SC, Adusumalli S, Asch DA, Groeneveld PW, Ungar LH, et al. Predicting cardiovascular risk using social media data: performance evaluation of machine-learning models. *JMIR Cardio* 2021 Feb 19;5(1):e24473 [[FREE Full text](#)] [doi: [10.2196/24473](https://doi.org/10.2196/24473)] [Medline: [33605888](https://pubmed.ncbi.nlm.nih.gov/33605888/)]

14. Lundberg S, Allen P. A unified approach to interpreting model predictions. 2017 Dec Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4, 2017; Long Beach, CA, USA p. 4768-4777.
15. Magoulas G, Prentza A. Machine learning in medical applications. In: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; Sep 20, 2001:300-307.
16. Wang X, Sontag D, Wang F. Unsupervised learning of disease progression models. 2014 Presented at: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 2014; New York, NY, USA p. 85-94. [doi: [10.1145/2623330.2623754](https://doi.org/10.1145/2623330.2623754)]
17. Juarez-Orozco LE, Martinez-Manzanera O, Storti AE, Knuuti J. Machine learning in the evaluation of myocardial ischemia through nuclear cardiology. *Curr Cardiovasc Imaging Rep* 2019 Feb 9;12(2):1-8. [doi: [10.1007/s12410-019-9480-x](https://doi.org/10.1007/s12410-019-9480-x)]
18. Sardar P, Abbott JD, Kundu A, Aronow HD, Granada JF, Giri J. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. *JACC Cardiovasc Interv* 2019 Jul 22;12(14):1293-1303 [FREE Full text] [doi: [10.1016/j.jcin.2019.04.048](https://doi.org/10.1016/j.jcin.2019.04.048)] [Medline: [31320024](https://pubmed.ncbi.nlm.nih.gov/31320024/)]
19. Elul Y, Rosenberg AA, Schuster A, Bronstein AM, Yaniv Y. Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning-based ECG analysis. *Proc Natl Acad Sci U S A* 2021 Jun 15;118(24):e2020620118 [FREE Full text] [doi: [10.1073/pnas.2020620118](https://doi.org/10.1073/pnas.2020620118)] [Medline: [34099565](https://pubmed.ncbi.nlm.nih.gov/34099565/)]
20. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using F-FDG PET of the brain. *Radiology* 2019 Feb;290(2):456-464 [FREE Full text] [doi: [10.1148/radiol.2018180958](https://doi.org/10.1148/radiol.2018180958)] [Medline: [30398430](https://pubmed.ncbi.nlm.nih.gov/30398430/)]
21. Seah J, Tang J, Kitchen A, Gaillard F, Dixon A. Chest radiographs in congestive heart failure: visualizing neural network learning. *Radiology* 2019 Feb;290(2):514-522. [doi: [10.1148/radiol.2018180887](https://doi.org/10.1148/radiol.2018180887)] [Medline: [30398431](https://pubmed.ncbi.nlm.nih.gov/30398431/)]
22. Ohta Y, Yunaga H, Kitao S, Fukuda T, Ogawa T. Detection and classification of myocardial delayed enhancement patterns on MR images with deep neural networks: a feasibility study. *Radiol Artif Intell* 2019 May;1(3):e180061 [FREE Full text] [doi: [10.1148/ryai.2019180061](https://doi.org/10.1148/ryai.2019180061)] [Medline: [33937791](https://pubmed.ncbi.nlm.nih.gov/33937791/)]
23. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019 Jan;25(1):70-74. [doi: [10.1038/s41591-018-0240-2](https://doi.org/10.1038/s41591-018-0240-2)] [Medline: [30617318](https://pubmed.ncbi.nlm.nih.gov/30617318/)]
24. Kwon J, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *JAMA* 2018 Jul 03;320(13):e008678. [doi: [10.1161/jaha.118.008678](https://doi.org/10.1161/jaha.118.008678)]
25. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, et al. Improving risk prediction in heart failure using machine learning. *Eur J Heart Fail* 2020 Jan 12;22(1):139-147 [FREE Full text] [doi: [10.1002/ehfj.1628](https://doi.org/10.1002/ehfj.1628)] [Medline: [31721391](https://pubmed.ncbi.nlm.nih.gov/31721391/)]
26. Daghistani TA, Elshawi R, Sakr S, Ahmed AM, Al-Thwayee A, Al-Mallah MH. Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. *Int J Cardiol* 2019 Aug 01;288:140-147. [doi: [10.1016/j.ijcard.2019.01.046](https://doi.org/10.1016/j.ijcard.2019.01.046)] [Medline: [30685103](https://pubmed.ncbi.nlm.nih.gov/30685103/)]
27. Gjoreski M, Gradišek A, Gams M, Simjanoska M, Peterlin A, Poglajen G. Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers. : *IEEE*; 2017 Aug 21 Presented at: 13th International Conference on Intelligent Environments; 2017; Seoul, South Korea p. 14. [doi: [10.1109/ie.2017.19](https://doi.org/10.1109/ie.2017.19)]
28. Oguz C, Sen SK, Davis AR, Fu Y, O'Donnell CJ, Gibbons GH. Genotype-driven identification of a molecular network predictive of advanced coronary calcium in ClinSeq® and Framingham Heart Study cohorts. *BMC Syst Biol* 2017 Oct 26;11(1):99-114 [FREE Full text] [doi: [10.1186/s12918-017-0474-5](https://doi.org/10.1186/s12918-017-0474-5)] [Medline: [29073909](https://pubmed.ncbi.nlm.nih.gov/29073909/)]
29. Rowin E, Maron M. The role of cardiac MRI in the diagnosis and risk stratification of hypertrophic cardiomyopathy. *Arrhythmia Electrophysiol Rev Internet Radcliffe Cardiology* Dec 2016;1. [doi: [10.15420/aer.2016.13.3](https://doi.org/10.15420/aer.2016.13.3)]
30. Hoey ETD, Teoh JK, Das I, Ganeshan A, Simpson H, Watkin RW. The emerging role of cardiovascular MRI for risk stratification in hypertrophic cardiomyopathy. *Clin Radiol* 2014 Mar;69(3):221-230. [doi: [10.1016/j.crad.2013.11.012](https://doi.org/10.1016/j.crad.2013.11.012)] [Medline: [24412355](https://pubmed.ncbi.nlm.nih.gov/24412355/)]
31. Hiemstra YL, Debonnaire P, Bootsma M, Schaliij MJ, Bax JJ, Delgado V, et al. Prevalence and prognostic implications of right ventricular dysfunction in patients with hypertrophic cardiomyopathy. *Am J Cardiol* 2019 Aug 15;124(4):604-612. [doi: [10.1016/j.amjcard.2019.05.021](https://doi.org/10.1016/j.amjcard.2019.05.021)] [Medline: [31204037](https://pubmed.ncbi.nlm.nih.gov/31204037/)]
32. Just E. Understanding Risk Stratification, Comorbidities, and the Future of Healthcare. 2014. URL: <https://www.slideshare.net/healthcatalyst1/understanding-risk-stratification-comorbidities-and-the-future-of-healthcare> [accessed 2015-06-05]
33. Hess OM. Risk stratification in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2003 Sep;42(5):880-881. [doi: [10.1016/s0735-1097\(03\)00838-6](https://doi.org/10.1016/s0735-1097(03)00838-6)]
34. Steriotis A, Sharma S. Risk stratification in hypertrophic cardiomyopathy. *Eur Cardiol* 2015 Jul;10(1):31-36 [FREE Full text] [doi: [10.15420/ecr.2015.10.01.31](https://doi.org/10.15420/ecr.2015.10.01.31)] [Medline: [30310420](https://pubmed.ncbi.nlm.nih.gov/30310420/)]
35. Maron BJ, Rowin EJ, Casey SA, Haas TS, Chan RH, Udelson JE, et al. Risk stratification and outcome of patients with hypertrophic cardiomyopathy ≥60 years of age. *Circulation* 2013 Feb 05;127(5):585-593. [doi: [10.1161/circulationaha.112.136085](https://doi.org/10.1161/circulationaha.112.136085)]
36. Maron MS, Rowin EJ, Wessler BS, Mooney PJ, Fatima A, Patel P, et al. Enhanced American College of Cardiology/American Heart Association strategy for prevention of sudden cardiac death in high-risk patients with hypertrophic cardiomyopathy. *JAMA Cardiol* 2019 Jul 01;4(7):644-657 [FREE Full text] [doi: [10.1001/jamacardio.2019.1391](https://doi.org/10.1001/jamacardio.2019.1391)] [Medline: [31116360](https://pubmed.ncbi.nlm.nih.gov/31116360/)]



37. Gersh B, Maron B, Bonow R, Dearani J, Fifer M, Link M, American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography/Interventions, Society of Thoracic Surgeons. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* 2011 Dec 13;124(24):e783-e831. [doi: [10.1161/CIR.0b013e318223e2bd](https://doi.org/10.1161/CIR.0b013e318223e2bd)] [Medline: [22068434](https://pubmed.ncbi.nlm.nih.gov/22068434/)]
38. Robnik-Šikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression. 1997 Presented at: Proceedings of the Fourteenth International Conference on Machine Learning (ICML) 1997; 1997; Nashville, TN, USA p. 296-304.
39. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJ, Population Health Metrics Research Consortium (PHMRC). Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr* 2011 Aug 04;9:29-11 [FREE Full text] [doi: [10.1186/1478-7954-9-29](https://doi.org/10.1186/1478-7954-9-29)] [Medline: [21816105](https://pubmed.ncbi.nlm.nih.gov/21816105/)]
40. Malley JD, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability machines: consistent probability estimation using nonparametric learning machines. *Methods Inf Med* 2012;51(1):74-81 [FREE Full text] [doi: [10.3414/ME00-01-0052](https://doi.org/10.3414/ME00-01-0052)] [Medline: [21915433](https://pubmed.ncbi.nlm.nih.gov/21915433/)]
41. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* 2014 Feb 21;8:14 [FREE Full text] [doi: [10.3389/fninf.2014.00014](https://doi.org/10.3389/fninf.2014.00014)] [Medline: [24600388](https://pubmed.ncbi.nlm.nih.gov/24600388/)]
42. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. USA: Association for Computing Machinery; 2016 Presented at: Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '16; Aug 13, 2016; San Francisco, CA, USA p. 785-794. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
43. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L. keras-team/keras-tuner: Hyperparameter Tuning for Humans. URL: <https://github.com/keras-team/keras-tuner> [accessed 2021-03-05]
44. Chen S, Bortsova G, Juárez A, van TG, de BM. Multi-task attention-based semi-supervised learning for medical image segmentation. *Lect Notes Comput Sci* 2019:457-465. [doi: [10.1007/978-3-030-32248-9\\_51](https://doi.org/10.1007/978-3-030-32248-9_51)]
45. Gu L, Zhang X, You S, Zhao S, Liu Z, Harada T. Semi-supervised learning in medical images through graph-embedded random forest. *Front Neuroinform* 2020 Nov 10;14:601829 [FREE Full text] [doi: [10.3389/fninf.2020.601829](https://doi.org/10.3389/fninf.2020.601829)] [Medline: [33240071](https://pubmed.ncbi.nlm.nih.gov/33240071/)]
46. Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G. Semi-supervised learning for network-based cardiac MR image segmentation. In: *Lecture Notes in Computer Science*. Cham: Springer; 2017:253-260.
47. Viceconti M, Henney A, Morley-Fletcher E. In silico clinical trials: how computer simulation will transform the biomedical industry. *Int J Clin Trials* 2016 May 09;3(2):37. [doi: [10.18203/2349-3259.ijct20161408](https://doi.org/10.18203/2349-3259.ijct20161408)]
48. Niederer SA, Lumens J, Trayanova NA. Computational models in cardiology. *Nat Rev Cardiol* 2019 Feb;16(2):100-111 [FREE Full text] [doi: [10.1038/s41569-018-0104-y](https://doi.org/10.1038/s41569-018-0104-y)] [Medline: [30361497](https://pubmed.ncbi.nlm.nih.gov/30361497/)]
49. Tannenbaum SJ, Holford NHG, Lee H, Peck CC, Mould DR. Simulation of correlated continuous and categorical variables using a single multivariate distribution. *J Pharmacokinet Pharmacodyn* 2006 Dec;33(6):773-794. [doi: [10.1007/s10928-006-9033-1](https://doi.org/10.1007/s10928-006-9033-1)] [Medline: [17053984](https://pubmed.ncbi.nlm.nih.gov/17053984/)]
50. Robnik Šikonja M. Dataset comparison workflows. *IJDS* 2018;3(2):126-145. [doi: [10.1504/ijds.2018.10013385](https://doi.org/10.1504/ijds.2018.10013385)]
51. Bøtcher SG, Dethlefsen C. DEAL: a package for learning Bayesian networks. In: *J Stat Softw*. Aalborg University: Department of Mathematical Sciences; 2003.
52. D'Agostino R, Stephens M. Goodness-of-Fit Techniques. New York, NY, USA: Marcel Dekker; 1986.
53. Robnik-Šikonja M, Bohanec M. Perturbation-Based Explanations of Prediction Models. Cham: Springer; 2018:159-175.
54. SHaRe. URL: <https://theshareregistry.org/home/patient> [accessed 2021-05-14]

## Abbreviations

- AI:** artificial intelligence
- BSA:** body surface area
- CMR:** cardiac magnetic resonance
- COPD:** chronic obstructive pulmonary disease
- ECG:** electrocardiogram
- Echo:** echocardiogram
- GB:** gradient boosted
- HCM:** hypertrophic cardiomyopathy
- KNN:** k-nearest neighbors
- LA\_d:** left atrial diameter
- LA\_Vol:** left atrial volume
- LR:** linear regression
- LVEF:** left ventricular ejection fraction LVIDd: left ventricular internal diameter at end diastole
- LVIDs:** left ventricular internal diameter at end systole

**MAE:** mean absolute error  
**ML:** machine learning  
**MVND:** multivariate normal distribution  
**NN:** neural network  
**NYHA:** New York Heart Association  
**RF:** random forest  
**RMSE:** root-mean-square error  
**RRMSE:** relative root-mean-square error  
**SCD:** sudden cardiac death  
**SHAP:** Shapley additive explanation  
**SVM:** support vector machine

*Edited by C Lovis; submitted 18.05.21; peer-reviewed by T Kahlon, P Li, P Bhadra; comments to author 27.09.21; revised version received 27.10.21; accepted 04.12.21; published 02.02.22*

*Please cite as:*

*Pičulin M, Smole T, Žunkovič B, Kokalj E, Robnik-Šikonja M, Kukar M, Fotiadis DI, Pezoulas VC, Tachos NS, Barlocco F, Mazzarotto F, Popović D, Maier LS, Velicki L, Olivotto I, MacGowan GA, Jakovljević DG, Filipović N, Bosnić Z*

*Disease Progression of Hypertrophic Cardiomyopathy: Modeling Using Machine Learning*

*JMIR Med Inform 2022;10(2):e30483*

*URL: <https://medinform.jmir.org/2022/2/e30483>*

*doi: [10.2196/30483](https://doi.org/10.2196/30483)*

*PMID:*

©Matej Pičulin, Tim Smole, Bojan Žunkovič, Enja Kokalj, Marko Robnik-Šikonja, Matjaž Kukar, Dimitrios I Fotiadis, Vasileios C Pezoulas, Nikolaos S Tachos, Fausto Barlocco, Francesco Mazzarotto, Dejana Popović, Lars S Maier, Lazar Velicki, Iacopo Olivotto, Guy A MacGowan, Djordje G Jakovljević, Nenad Filipović, Zoran Bosnić. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 02.02.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.