Disease Progression of Hypertrophic Cardiomyopathy: Modeling Using Machine Learning

Disease progression of hypertrophic cardiomyopathy: Modeling using machine learning

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

Submitted to: JMIR Medical Informatics
on: May 18, 2021

Table of Contents

Disease progression of hypertrophic cardiomyopathy: Modeling using machine learning

Matej Pi?ulin[1*] PhD; Tim Smole[1*] MSc; Bojan Žunkovi?[1*] PhD; Enja Kokalj[1*] MSc; Marko Robnik-Šikonja[1*] PhD, Prof Dr; Matjaž Kukar[1*] PhD, Prof Dr; Dimitrios I Fotiadis[2*] PhD, Prof Dr; Vasileios C Pezoulas[2*] MSc; Nikolaos S Tachos[2*] MSc; Fausto Barlocco[3*] PhD, MD; Francesco Mazzarotto[3*] PhD, MD; Dejana Popovi?[4*] PhD, MD, Prof Dr; Lars S Maier[5*] PhD, MD, Prof Dr; Lazar Velicki[6, 7*] PhD, MD, Prof Dr; Iacopo Olivotto[3*] PhD, MD; Guy A MacGowan[8*] PhD, MD; Djordje G Jakovljevi?[8, 9*] PhD, MD, Prof Dr; Nenad Filipovi?[10*] PhD, Prof Dr; Zoran Bosni?[1*] PhD, Prof Dr

[1]Faculty of Computer and Information Science University of Ljubljana Ljubljana SI
[2]Unit of Medical Technology and Intelligent Information Systems Dept. of Materials Science and Engineering University of Ioannina Ioannina GR
[3]Cardiomyopathy Unit Careggi University Hospital University of Florence Florence IT
[4]Clinic for Cardiology Clinical Center of Serbia University of Belgrade Belgrade RS
[5]Dept. of Internal Medicine II (Cardiology, Pneumology, Intensive Care Medicine) University Hospital Regensburg Regensburg DE
[6]Faculty of Medicine University of Novi Sad Novi Sad RS
[7]Institute of cardiovascular diseases Vojvodina Sremska Kamenica RS
[8]Translational and Clinical Research Institute Faculty of Medical Sciences Newcastle University Newcastle upon Tyne GB
[9]Faculty of Health and Life Sciences Coventry University Coventry GB
[10]BIOIRC - Bioengineering Research and Development Center Kragujevac RS
[*]these authors contributed equally

Corresponding Author:
Matej Pi?ulin PhD
Faculty of Computer and Information Science
University of Ljubljana
Ve?na pot 113
Ljubljana
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 out 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 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 ML-based tool for predicting disease progression for patients diagnosed with HCM in terms of adverse remodeling of heart during a 10-year period.

Methods: The method consists of six predictive regression models that independently predict future values of six clinical characteristics: left atrial size, left atrial volume, left ventricular ejection fraction, New York Heart Association Functional Classification (NYHA), left ventricular internal diastolic diameter, and left ventricular internal systolic diameter. We supplemented each prediction with the explanation that is generated with the Shapely additive explanation (SHAP) method.

Results: The final experiments show that predictive error is lower on 5 out of 6 constructed models with comparison to experts or consortium of experts. The experiments revealed that semi-supervised learning and the artificial data from virtual patients helped to achieve even higher predictive accuracies.

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

(JMIR Preprints 18/05/2021:30483)
DOI: https://doi.org/10.2196/preprints.30483

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Authors: Matej Pičulin[1], PhD; Tim Smole[1], MSc; Bojan Žunkovič[1], PhD; Enja Kokalj[1] , MSc; Marko Robnik-Šikonja[1], prof., PhD; Matjaž Kukar[1], prof., PhD; Dimitrios I Fotiadis[2], prof., PhD; Vasileios C Pezoulas[2], MSc; Nikolaos S Tachos[1] ,MSc; Fausto Barlocco[3], PhD, MD; Francesco Mazzarotto[3], PhD, MD; Dejana Popović[4] , prof., PhD, MD; Lars S Maier[5], prof., PhD, MD; Lazar Velicki[6], prof., PhD, MD; Iacopo Olivotto[3], PhD, MD; Guy A MacGowan [7], PhD, MD; Djordje G Jakovljević[7,8], prof., PhD, MD; Nenad Filipović [9], prof., PhD; Zoran Bosnić[1], prof., PhD

Corresponding author:
Matej Pičulin, PhD
Faculty of Computer and Information Science
Večna pot 113,
1000 Ljubljana,
Slovenia
Phone: +386 1 479 8226
Email: matej.piculin@fri.uni-lj.si

Affiliations:
[1] University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, Ljubljana, Slovenia
[2] University of Ioannina, Dept. of Materials Science and Engineering, Unit of Medical Technology and Intelligent Information Systems
[3] Cardiomyopathy Unit, Careggi University Hospital, University of Florence, Italy
[4] University of Belgrade, Clinic for Cardiology, Clinical Center of Serbia, Faculty of Pharmacy, Belgrade, Serbia
[5] University Hospital Regensburg, Dept. of Internal Medicine II (Cardiology, Pneumology, Intensive Care Medicine), Germany
[6] Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia and Institute of cardiovascular diseases Vojvodina, Sremska Kamenica, Serbia
[7] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[8] Faculty of Health and Life Sciences, Coventry University, Coventry, UK
[9] BIOIRC - Bioengineering Research and Development Center, Kragujevac, Serbia

Disease progression of hypertrophic cardiomyopathy: Modeling using machine learning

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 out of 500 young adults and can cause sudden cardiac death (SCD).
Objectives: Although the current state-of-the-art methods model the risk of SCD for patients, to 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 ML-based tool for predicting disease progression for patients diagnosed with HCM in terms of adverse remodeling of heart during a 10-year period.

Methods: The method consists of six predictive regression models that independently predict future values of six clinical characteristics: left atrial size, left atrial volume, left ventricular ejection fraction, New York Heart Association Functional Classification (NYHA), left ventricular internal diastolic diameter, and left ventricular internal systolic diameter. We supplemented each prediction with the explanation that is generated with the Shapely additive explanation (SHAP) method.
Results: The final experiments show that predictive error is lower on 5 out of 6 constructed models with comparison to experts (on the average by 0.34) or consortium of experts (on the average by 0.22). The experiments revealed that semi-supervised learning and the artificial data from virtual patients helped to 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 performance of experts for five out of six targets.

Keywords: hypertrophic cardiomyopathy; disease progression; machine learning; artificial intelligence

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 is becoming 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’s disease [3,4], Parkinson’s 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, which 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 pre-emptive prediction, prevention and proactive management, thus this opens an opportunity for methods of artificial intelligence [2]. Disease progression is especially hard to detect in slow progressing diseases, such as the HCM that is present in about 1 out of 500 young adults [10]. Although HCM has four identified stages [11], HCM patients can experience a sudden cardiac arrest or the disease can progress slowly 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 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 modeled recently for atherosclerotic cardiovascular disease [12].

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 heart during a 10-year period. The method consists of six contemporaneous predictive regression models that independently predict future values of the following six clinical characteristics: left atrial size (LA), left atrial volume (LA_Vol), left ventricular ejection fraction (LVEF), New York Heart Association Functional Classification (NYHA), left ventricular internal diastolic diameter (LVIDd), and left ventricular internal systolic diameter (LVIDs). Each prediction is supplemented with the explanation that is generated with the Shapely additive explanation (SHAP) method [14]. Comparison between current and future values of these six 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 becoming frequently applied in medicine to improve prediction of disease progression, extraction of medical knowledge for outcome research, therapy planning and support, and for the overall patient management [15]. A wide variety of ML approaches turned out to solve challenging problems in these tasks. For example, diseases such as Alzheimer's disease, Diabetes, chronic obstructive pulmonary disease (COPD) progress slowly over the years. For modeling of COPD, a Markov model was proposed by [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.

Further, a hybrid approach for progression of Parkinson’s disease [5] was successfully used by combining a variety of machine learning methods from different families: clustering, dimensionally reduction and incremental support vector regression. Deep learning was used for predicting the Alzheimer's disease on average about six years in advance [20] and for modeling Alzheimer’s 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 machine learning and appropriate data.

In cardiology, there were several works addressing disease progression trends related to different cardiological diseases. With the increase of computational power, machine ML has become a tool to analyze non-linear dependencies that are present either in relational data or images. Juarez-Orozco et al. [17] emphasized the advantages of ML, especially of deep learning, in cardiac nuclear imaging, where ML can aid with ischemia diagnosis and event prognosis. Sardar et al. [18] emphasize the advantages of artificial intelligence 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 ECG analysis, Elul et al. [19] also state the crucial disadvantages of ML, which include lack of explanation, relating the automated diagnosis with medical knowledge, and transparency of system’s limitations. In their work, the authors proceed to flagging individual predictions, which are irrelevant or not useful. To summarize, the mentioned works characterize the AI as a developing tool that, with the synergy between man and machine, which help transform medical practice and clinical care.

Several other ML approaches also model disease progression well in other medical domains, such as the kidney disease progression [7]. In this work, nine ML approaches were tested: linear regression, elastic net regression, lasso regression, ridge regression, support vector machines (SVM), random forests, k-nearest neighbors (KNN), neural networks, 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 to 10 years, such as KNN, decision trees, linear regression, and SVM. Support vector machines performed best.

To summarize, the above overview indicates that the 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 were 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 focuses on risk stratification of patients. One of the biggest obstacles for utilizing data for broader variety of ML applications is that data are usually stored in diverse repositories which are not readily utilizable for cardiovascular research due to various data quality challenges [2]. Where the data was readily available, different ML algorithms have been successfully used, such as the Wasserstein generative adversarial networks (GAN) [21], convolutional neural networks [22,23], deep neural networks [24], and boosted decision trees [25]. Some authors tested multiple models, such as random forests, artificial neural networks, SVM and Bayesian networks [26], or a combination of J48, naïve Bayes, KNN, SVM, random forest, 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 analyzing image data. For the latter problems, the most successful methods were neural networks, deep neural networks and convolutional networks. Advances in prediction accuracy have also been made by using deep neural networks to make predictions based on fast large-scale genome-wide association studies [28].

HCM is a severe disease for which four stages of its progression have been identified in medical literature [11]. The current ML state-of-the-art mostly utilizes only statistical models, such as the multivariate regression analysis, which utilize pre-selected predictor variables of known medical importance. Cardiac magnetic resonance (CMR) images [29,30] and echocardiographic diagnostics [31] were found to be a good source of important attributes for HCM identification. Recently, researchers started proposing ML-based risk stratification for patients diagnosed with to separate patients into low- and high-risk categories or several categories on a scale [32]. Medical literature is mostly focused on finding risk factors that identify increased risk of SCD in patients with HCM [12,33]. A study [34] presented the guidelines used in risk stratification for patients with HCM proposed the potential SCD modifiers. Maron et al. [35] performed a similar study on older population 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 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 the HCM risk. In our work, we focus on modeling the future development of HCM by predicting the change in relevant cardiac parameters for 10-years ahead.

Aims and contributions

Novelties and contributions of this paper include:

- disease progression system that comprises models for prediction of six 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 is extended using semi-supervised learning or with artificial patient data,
- validation of predictive accuracy with medical experts by comparing machine learning 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 a slow progression of HCM based on longitudinal data.

Methods

Modeling disease progression

In this work, we model the disease progression by predicting six relevant patients' 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 Section Dataset. The proposed disease progression system (Figure 1) takes as an input patients' clinical data and data about their past disease-related events, such as dates of atrial fibrillation or syncope. The output of the system is a set of six contemporaneous target predictions for parameters:

- Left atrial diameter (LA),
- Left atrial volume (LA_Vol),
- Left Ventricular Ejection Fraction (LVEF),
- Left Ventricular Internal Dimension at end-Diastole (LVIDd),
- Left Ventricular Internal Dimension at end-Systole (LVIDs), and
- New York Heart Association (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 of the six target variables throughout the 10-year period.

We train the proposed disease progression system using supervised ML techniques. To further improve the results, we augment the original data using unlabeled data (semi-supervised learning) and virtual patients' data. We apply the semi-supervised learning using patients without 10-year follow-ups and generate virtual patients' data using various techniques for artificial data generation. The semi-supervised learning first predicts patients' targets using the trained models on labeled data, so they can be afterwards included into the training dataset. In the following subsections, we describe the dataset, predictive modeling with supervised models, use of semi-supervised learning and virtual patient data, and generation of prediction explanations.

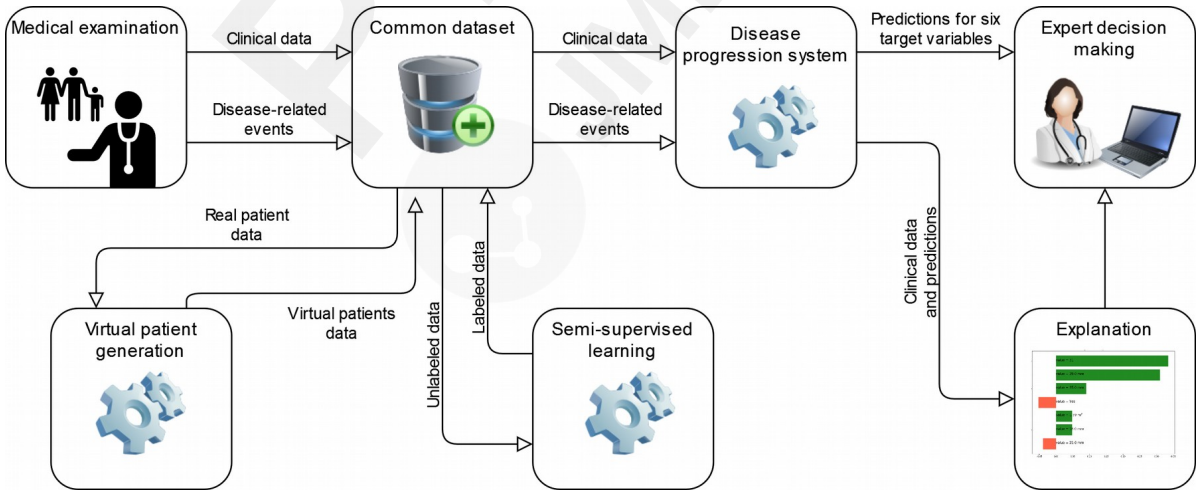


Figure 1: An overview of the proposed disease progression system. The system receives clinical data and disease-related events of a patient as an input, utilizes virtual patient data and semi-supervised learning for self-improvement, and returns the predictions and their explanation for six target variables.

Dataset

The proposed approach was developed on a dataset that was provided by the University of Florence as a result of their long-term clinical practice. The dataset included patients who were enrolled over the last 40 years (Figure 2), 75 % of them after the year 2000. They were followed for an average duration of about 7 years and ranging up to 37 years. The dataset contains longitudinal clinical data for 2,318 patients diagnosed with HCM or patients that had an HCM diagnosed relative (1,457 male and 861 female patients). During the patients' visits, various clinical tests and relevant disease-related events were recorded. These data include: general data (gender, age, height, weight etc.), genetic data (detected mutations), clinical tests (echo, Holter monitory, blood test, CMR, stress test), prescribed medications (type, start date, termination date), and disease-related events (e.g. SCD, heart failure, transplant, abnormal Holter, pacemaker or implantable cardioverter defibrillator (ICD) implantation etc.). 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 employed selectively due to its greater accuracy in measuring volumes. Although echo and CMR were treated separately and never computationally compared to each other in medical practice, we use the CMR, where available, as an additional data modality to possibly improve the prediction accuracy. In total, there were 6,227 events recorded, out of which 4,902 events occurred to patients who were primarily diagnosed with HCM. The structure of the dataset therefore allows observing how patients' clinical characteristics change over time, which is essential for the desired modeling of HCM progression. The basic patient characteristics are shown below in (Table 1) for continuous parameters, (Table 2) for binary parameters and (Table 3) for the remaining parameters. The characteristics are extracted from 10,318 measurements made in total. Additionally, (Table 4) also shows the missing data numbers and percentages for the six selected target variables for their role as input or target variables.

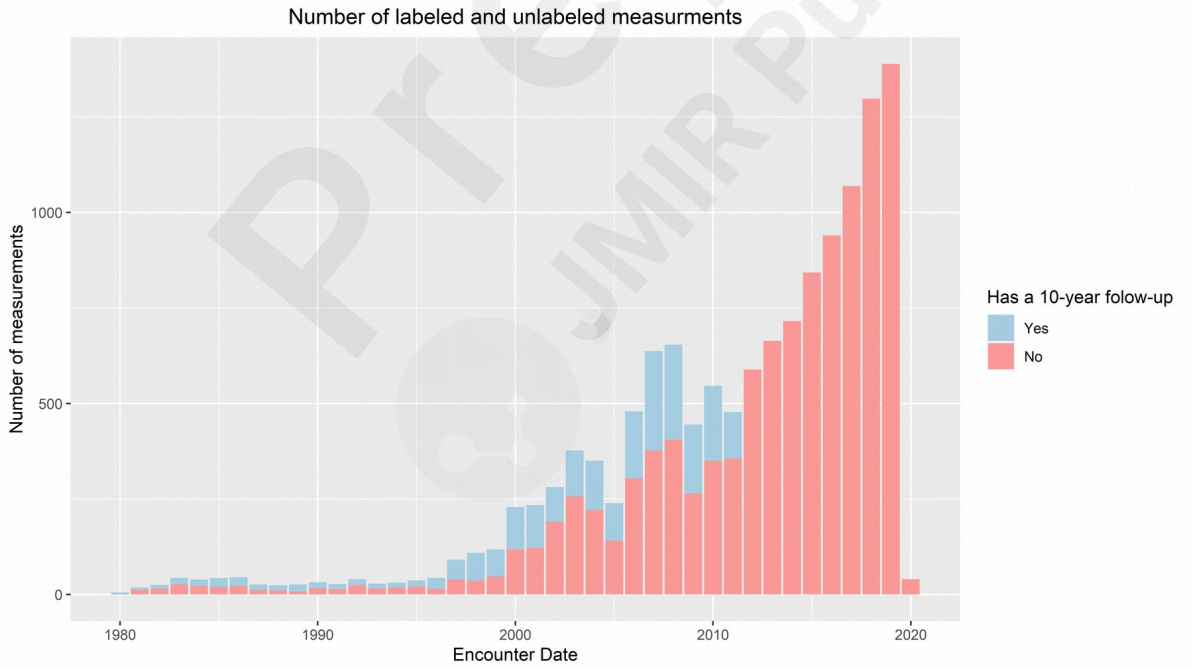


Figure 2: This figure shows relationship between the amount of labeled and unlabeled data. The bars for Yes and No values are stacked, visually revealing the ratio between the labeled and unlabeled data. Note that the rightmost columns do not have 10-year follow up data, as they are younger than 10 years.

Table 1: Basic characteristics of the patients for basic continuous parameters.

Continuous parameter	Mean	Standard deviation	Missing % (#)
Age	52.1 (years)	18.6	0.04 (4)
Weight	73.4 (kg)	14.6	23.1(2381)
Height	169 (cm)	10.3	22.0(2273)
BMI	25.6	4.09	23.5(2423)
NYHA	1.69	0.73	9.53(983)

Table 2: Basic characteristics of the patients for basic binary parameters.

Binary parameters	1-value % (#)	0-value % (#)	Missing % (#)
Alcohol	1.0 (yes) (103)	99.0 (no) (10215)	0.0 (0)
Drug	0.2 (yes) (18)	99.8 (no) (10300)	0.0 (0)
Smoking	33.3 (yes) (3437)	66.7 (no) (6881)	0.0 (0)
Pregnancy	4.3 (yes) (443)	95.7 (no) (9875)	24.4 (2515)
Gender	62.0 (male) (6400)	38.0 (female) (3918)	0.0 (0)

Table 3: Basic characteristics for groups of parameters.[a]

Aggregated	#parameters	Missing % avg (#)
ECG	9	49.4 (45839)
Echo	26	36.6 (98191)
CMR	10	78.7 (81174)

[a]*The table shows aggregated statistics for several parameters obtained from the same procedure (CMR = Cardiovascular magnetic resonance imaging, ECG = electrocardiogram, Echo = echocardiogram).*

Table 4: *Percent and absolute number of missing values of target variables as class and as input.*

	LA	LVEF	NYHA	LVIDd	LVIDs	LA_Vol
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Target						
	8569 (83.0)	8481 (82.2)	8313 (80.6)	8607 (83.4)	9336 (90.5)	8631 (83.6)
Input						
	2691 (26.1)	2399 (23.3)	983 (9.5)	2517 (24.4)	5329 (51.6)	3680 (35.7)

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

1. **Forming training examples**: Since not all clinical tests can be done on the same day or

month, we defined a training example as a set of measurements within a time frame of one 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 progresses slowly. If the patient had a certain test performed multiple times within this time frame, multiple tests were treated as separate measurements. In case that a certain type of test was not performed in the one-year time frame, the corresponding variables were recorded as missing. Constructing training examples in this way yields a dataset with 13,386 examples, with 3.9 ± 4.8 examples per patient.

2. **Imputation of missing data**: The missing values in the dataset, either because of non-performed test or the erroneous input of a datum, were imputed either by: copying closest past values (sensible because the progression of HCM is slow; used on numerical and categorical attributes), by imputing values of a healthy patient (sampled from normal distribution; used for numerical attributes), or by 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 an 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 neural networks, we also scaled the values to the interval [0, 1] and encoded nominal values using the one-hot-encoding for.

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

Table 5: Selected attributes using RReliefF.[a]

Variable name	LA	LVEF	NYHA	LVIDd	LVIDs	LA_Vol	avg rank
Anthropometric parameters							
Age	0.198	0.194	0.166	0.142	0.166	0.158	1.000
Gender	0.051	0.037	0.043	0.055	0.058	0.022	12.500
Height	0.057	0.064	0.045	0.075	0.051	0.029	9.167
BSA	0.075	0.073	0.053	0.095	0.085	0.045	4.167
Risk factors							
Smoking	0.063	0.046	0.052	0.032	0.069	0.082	7.500
Presence of hypercholesterolaemia	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
Family history of HCM	0.056	0.060	0.061	0.047	0.052	0.066	5.833
Family history of SCD	0.027	0.051	0.032	0.031	0.051	0.049	14.667
Clinical, ECG and Echo parameters							
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
	IVS	0.043	0.052	0.049	0.041	0.057	0.052	8.167
	LA	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
Genetics								
	Mutation MYBPC3	0.045	0.041	0.039	0.051	0.052	0.059	9.667
	Mutation MYH7	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

[a]The table shows RReliefF feature scores and the average ranks for each target variable. Names of ten highest ranked variables are underlined. (BSA – body surface area, HCM – hypertrophic cardiomyopathy, IVS - interventricular septum, SCD – sudden cardiac death, LA – left atrium, LA_Vol – left atrium volume, LVIDs - left ventricular internal dimension in systole, LVIDd - left ventricular internal dimension in diastole, LVEF – left ventricular ejection fraction, NYHA - New York Heart Association)

Predictive modeling with supervised and semi-supervised machine learning

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

- **Random forests** [39,40] is an ensemble prediction model that constructs multiple randomized decision trees. The implementations of random forest classifier in statistical package R (library *ranger*) and 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]: an ensemble of weak decision tree predictors, implemented in the open source software library XGBoost;
- **Linear regression:** a traditional method of finding a linear dependence between attributes and the selected target variable;
- **Neural networks** mimic the architecture and working of brain neurons. We used one input and one output layer and one or several hidden layers. In the optimization process, we optimized several learning parameters, such as learning rate, number of hidden layers, sizes of layers, regularization, sample weights, class weights, dropout, and batch normalization.

The best hyper parameters of the above algorithms were tuned using the Bayesian optimization and random search implemented in *keras-tuner* [43].

Semi-supervised learning and virtual patients

Semi-supervised learning is increasingly used in medicine especially for the 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 (i.e., 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 semi-supervised learning, producing a teacher model. The unlabeled examples were labeled with the supervised learning predictive model (see previous 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 as the teacher model.

To further improve the results of semi-supervised learning, we utilized artificially generated data

(i.e., virtual patients). Virtual data generation can sometimes replace experiments in biomedical experiments on animals [47]. Specifically in cardiovascular modeling patient-specific virtual patient modeling 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, log-MVND) [49], and non-parametric methods using supervised tree ensembles, unsupervised tree ensembles, RBF-based neural networks [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 Dataset section.

Explanation of the predictive model

Supervised ML models often exhibit a black-box nature, meaning that they can model data but not provide the explanation for the contained knowledge as well as the reasoning used in predictions. This means that the model lacks transparency and interpretability. To address this, explanation methods provide justification for each prediction and assess features with the highest impact [53]. This is very 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 explanation method SHAP (SHapley Additive exPlanations) [14] that is model-agnostic method, generating 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 contributed 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 model’s predictions.

Results

Models’ comparison

To evaluate and compare the performance of six predictive models, we used stratified 10-fold cross-validation. For each of the six predictive problems, four different regression models were evaluated (linear regression - LR, random forest - RF, gradient boosted trees – GB, and neural networks - NN). The following parameters were varied in tests:

- application of semi-supervised learning (denoted with S),
- adding virtual patients' data into the learning dataset (denoted with VP),
- use of all 112 features (denoted with All) or only the subset of 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 above parameters were used in experiments. Some combinations were omitted due to limitations (e.g. VP generators cannot generate data for all 112 attributes, so VP was evaluated only with the subset of attributes) or excessive time complexity (e.g., the use of virtual patients with neural networks).

Performance of predictive models

To compare the accuracy of the obtained models, we computed the following four metrics: mean absolute error (MAE), root mean squared error (RMSE), and two variations of the relative root mean squared error ($RRMSE_{mean}$ and $RRMSE_{const}$). Mean absolute error (MAE) measures the average absolute difference between prediction and true value over all examples in the test set. Root mean squared error (RMSE) addresses the issue that the squared values of MSE are hard to interpret. RRMSE measures a relative ratio between the obtained model and the baseline model. We computed two variations of the RRMSE with two different baseline models: mean predictor and constant predictor. With $RRMSE_{mean}$ we compared the performance of the obtained model to the model that returns 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 assumes that the value of the target variable will remain constant/unchanged over the 10-year period (constant predictor).

We summarized (Table 6) the performance of the best preforming predictive models (RF, LR, GB, NN) and parameters (S, VP, all/subset) for each target variable. We can see that top performing regression models are RF and GB for all target variables. We achieved the best results by applying semi-supervised learning (S) for all target variables and using virtual patients (VP) for five out of six target variables. For all targets, the best results were obtained by learning from a subset of 21 most important features. The values of both RRMSE metrics reveal that the model perform better than the baseline models (their values are lower than 1.0), with model for the LA target achieving the lowest predictive error.

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

Target	**Model and parameters**	**MAE**	**RMSE**	**$RRMSE_{mean}$**	**$RRMSE_{const}$**
LA	RF: S+VP+subset	3.4	4.73	0.54	0.46
LA_Vol	RF: S+VP+subset	18.4	26.73	0.56	0.47
LVEF	GB: S+subset	4.92	6.73	0.67	0.61
LVIDd	RF: S+VP+subset	3.53	5.26	0.68	0.64
LVIDs	RF: S+VP+subset	3.42	4.81	0.66	0.56
NYHA	RF: S+VP+subset	0.39	0.5	0.67	0.66

To further evaluate the contribution of different data augmentation strategies, we compared the results on different patients sets: original (all features), subset of best features, virtual patients (VP), semi-supervised learning (S), and the combination of the latter two (S + VP). The obtained results, shown for the best performing Random Forest model, are shown 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.

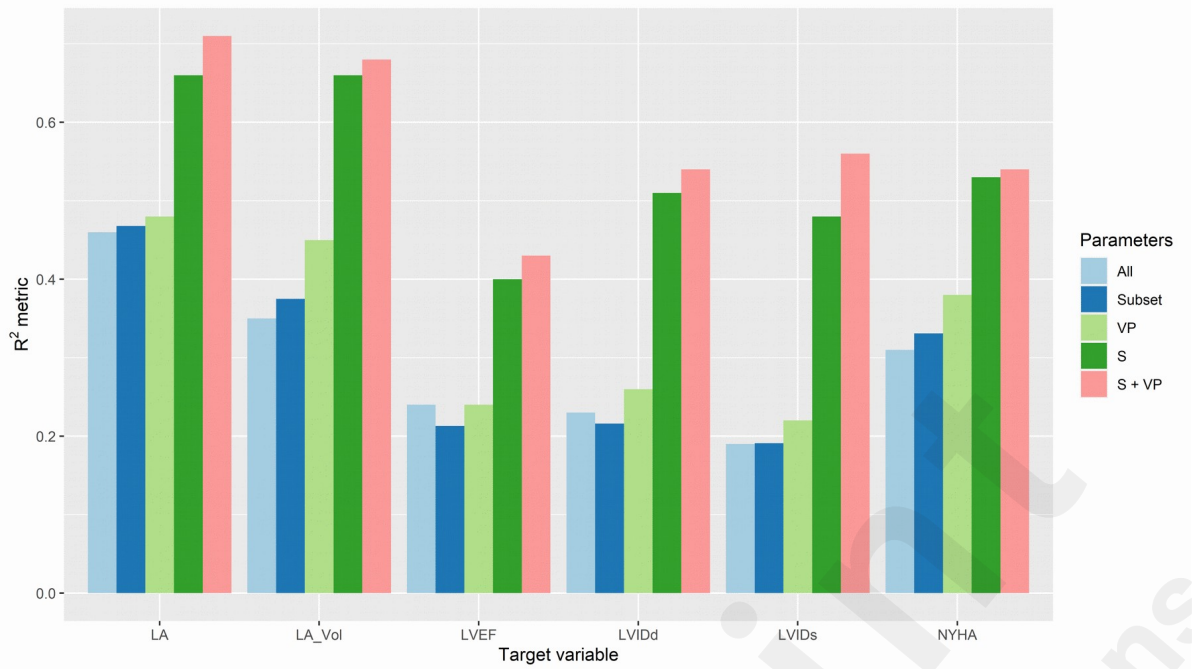


Figure 3: The plotted results for R^2 statistic for each target variable using different sets (input parameters). Note that VP, S and S + VP are used on feature subsets.

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

Explanation of predictions

To augment the output of prediction models, we applied the method SHAP (SHapley Additive exPlanations) [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 influenced the prediction along with their contribution values that define if the feature value either supports the predicted value or opposes it. The direction of the impact (i.e. sign of the contribution value) is denoted using different colors.

An example of generated explanation for the prediction for target LA (Figure 4) is presented below. Features' contributions are sorted in the 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 contributing factors to the increase or decrease of the LA value. We can see that the features "LA", "Atrial fibrillation", "Age", "Mutation MYBPC3", and "LVIDs" contributed to the increase of predicted value for LA over time, while "LA_Vol" and "Mutation MYH7" contributed to the decrease of predicted value for LA. 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 (46.00). Larger magnitudes of the features' contributions correspond to larger changes in the prediction value. For example, "LA" contributed the most (approximately 30%) to the increase in the predicted value.

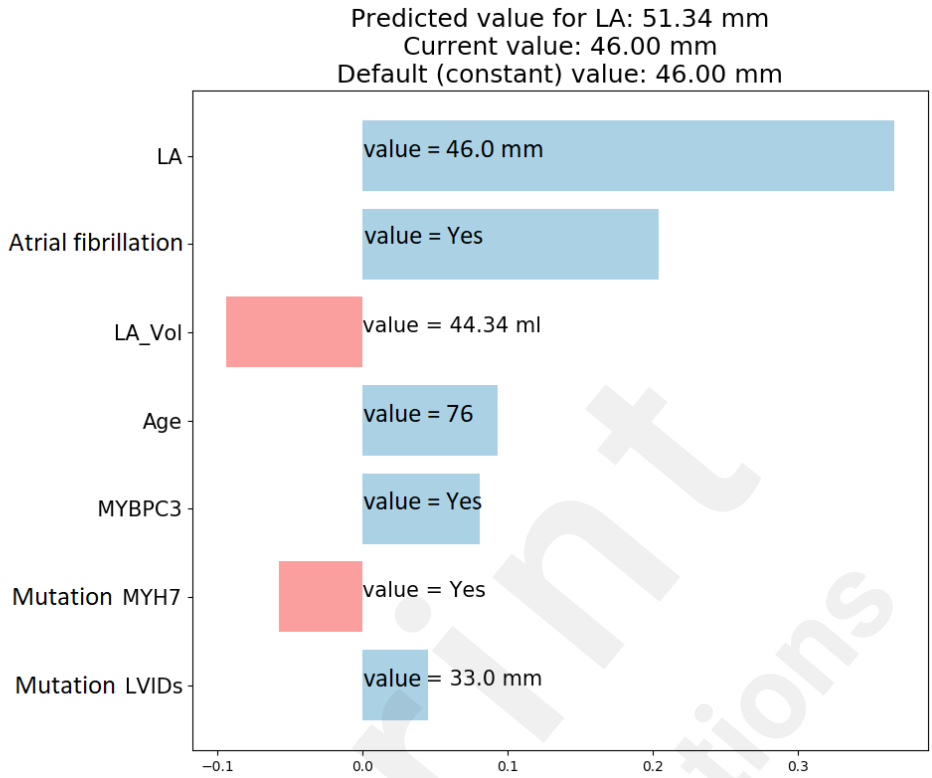


Figure 4: An example of explanation of prediction for target variable LA

Validation with medical experts

Besides evaluation of prediction models with statistical measures conducted in two previous sections, we engaged medical experts to provide further interpretation and validation of the results. First, we compare the accuracy of predictive models with the accuracy of human experts, which was obtained by using a survey (Multimedia Appendix 2). Second, we check if prediction explanations are sensible and consistent with 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 medications data) for ten patients, and the experts were asked to study them and complete the following two tasks:

1. to predict the magnitude of the 10-year change in the six studied clinical parameters (LA, 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;
2. to evaluate if the statements generated from the explanation (for example: "The current value of parameter “LA” will cause a decrease of the LA") are true or false. For each patient, six such statements were generated, covering the features with the highest contribution. More specifically, the questionnaire included evaluation questions for 6 parameters that contribute to change in LA, 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 ± 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 machine learning model, we first discretized the model's predictions into discrete intervals, so that they can be compared to the discrete intervals, predicted by the experts. We performed the discretization using bins of width 0.25σ, where σ was the standard deviation 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 expert (denoted with **E**),
- mean prediction error of the consortium prediction (i.e., the average prediction of all doctors') (denoted with **C**),

We can see that the mean prediction error of the discretized model MD (Table 7) is the lowest for all target variables except for LA. The mean errors of consortium predictions C are 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 and thus better than the error of the machine learning model.

Table 7: The mean absolute error of the discretized model predictions (MD), individual experts (E) and the entire consortium (C). The lowest achieved errors are denoted with italics.

Target/Prediction	Model (MD)	Expert (E)	Consortium (C)
NYHA	*0.30 ± 0.48*	0.84 ± 0.69	0.56 ± 0.34
LA	1.70 ± 0.82	1.69 ± 0.97	*1.66 ± 0.70*
LA_vol	*1.00 ± 0.82*	1.25 ± 0.98	1.13 ± 0.63
LVIDd	*0.80 ± 0.63*	1.09 ± 0.91	1.00 ± 0.77
LVIDs	*0.50 ± 0.71*	1.02 ± 0.86	0.88 ± 0.68
LVEF	*0.90 ± 0.88*	1.32 ± 0.90	1.28 ± 0.79

Validation of model explanation

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

The results (Table 8) of the analysis provide 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 that are shown in the decreasing order of the last column, show that the majority of the experts agree especially with the explanations for targets NYHA (average agreement of 0.73) and LVIDd (average agreement of 0.52). By comparing (Table 7) and (Table 8), we consistently see that the experts least agreed with explanations for target LA, for which the predictive model achieved 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 explanation is lower (Table 8) (e.g. for LVEF, LA_vol and LVIDs), there are three 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 explanation for parameters that contribute to predicting each target variable. Names of parameters with agreement higher than 50% are emphasized with italic. The last two columns provide *summary statistics.*

			Summary	
Target variable and parameters		**Expert agreement**	**Ratio of agreed features from at least 50% of experts**	**Average agreement**
NYHA				
	LA	*0.77*		
	Age	*0.77*	1.00 (4/4)	0.73
	LA_Vol	*0.62*		
	Atrial fibrillation	*0.77*		
LVIDd				
	BSA	0.15		
	Gender	*0.85*		
	LVIDd	*0.65*	0.67 (4/6)	0.52
	QRS duration	*0.69*		
	LVEF	0.23		
	Mutation MYH7	*0.54*		
LVEF				
	QRS duration	0.38		
	Presence of hypercholesterolaemia	*0.54*		
	Syncope	0.46	0.40 (2/5)	0.49
	Gene_Testing_Performed	*0.69*		
	NYHA	0.38		
LA_Vol				
	LA_Vol	*0.69*		
	BSA	*0.54*	0.75 (3/4)	0.48
	Age	0.15		
	Atrial fibrillation	*0.54*		
LVIDs				
	LA	0.38		
	LVIDd	0.38		

	LA_Vol	*0.62*		
	BSA	*0.85*	0.43 (3/7)	0.47
	Mutation MYBPC3	*0.62*		
	IVS	0.38		
	Family history of HCM	0.08		
LA				
	LA	*0.85*		
	Atrial fibrillation	0.15		
	BSA	0.08	0.17 (1/6)	0.36
	IVS	0.38		
	Age	0.31		
	LVEF	0.38		

Discussion

Principal Results

We presented a disease progression system for patients diagnosed with HCM that is based on predicting six target parameters (LA, LA_vol, LVIDd, LVIDs, LVEF, and NYHA) for 10 years ahead using supervised machine learning 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 semi-supervised learning and the artificial data from virtual patients helped to achieve even higher predictive accuracy for all six 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 five out of six targets.

Limitations

The design of the study carries several limitations, stemming from the fact that this work is based on real-world data that is expensive to obtain and is subject to noise. The first limitation of this study is that it is based only on a single medical center dataset. To further validate this study, it would be beneficial to independently evaluate the models with datasets from other centers or extend the existing dataset with more data. Additionally, the benefit for including more data could also be in diminishing a potential bias of our dataset, which could potentially include 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 unavailability of such data at the time of the study or data that was structured very differently, we leave this for our further work.

Further, to prepare the data to be used for machine learning and obtain stable predictions, we used several preprocessing and data augmentation steps. Since we are dealing with real medical data, this opens questions 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 timeframe and predicted risk timeframe influence the achieved accuracies. 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 opens several ideas forfuture study directions. First, we shall evaluate the proposed system on an independent cardiological dataset, e.g. The Sarcomeric Human Cardiomyopathy Registry – SHaRe [54]. Second, as our current approach provides future predictions for six independent parameters, the outputs shall 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 shall be tested (additional predictive models and feature selection techniques, including deep learning), as well as the reasons for the experts' disagreement with some of the explanation components shall be determined.

Conclusions

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 sudden cardiac death. Additionally, the results confirmed findings in [44–46] that additional artificial data and semi-supervised 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 [36].

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 777204 (www.silicofcm.eu) This article reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains.

Conflicts of Interest

None declared.

Abbreviations

AUC: area under curve
BSA: body surface area
CMR: cardiac magnetic resonance
COPD: chronic obstructive pulmonary disease
ECG: electrocardiogram
Echo: echocardiogram
GAN: generative adversarial network
GB: gradient boosted trees
HCM: hypertrophic cardiomyopathy
ICD: implantable cardioverter defibrillator
IVS: interventricular septum
KNN: k-nearest neighbors
LA: left atrial size
LA_Vol: left atrial volume
LR: linear regression
LVEF: left ventricular ejection fraction
LVIDd: left ventricular internal diastolic diameter

LVIDs: left ventricular internal systolic diameter
MAE: mean absolute error
ML: machine learning
MSE: mean squared error
MVND: multivariate normal distribution
NN: neural network
NYHA: New York heart association
RBF: radial basis function
RF: random forest
RMSE: root mean square error
RRMSE: relative root mean squared error
SCD: sudden cardiac death
SHAP: Shapely additive explanation
SHARE: sarcomere human cardiomyopathy registry
SVM: support vector machine

Multimedia Appendix 1

Bar graphs of parameter influence for each model used.

Multimedia Appendix 2

A sample of questionnaire for the first patient.

References

1. Cuocolo R, Perillo T, De Rosa E, Ugga L, Petretta M. Current applications of big data and machine learning in cardiology [Internet]. J Geriatr Cardiol. Science Press; 2019 [cited 2020 Dec 23]. p. 601–607. PMID:31555327
2. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: Are we there yet? [Internet]. Heart. BMJ Publishing Group; 2018 [cited 2020 Dec 23]. p. 1156–1164. PMID:29352006
3. Fisher CK, Smith AM, Walsh JR, Simon AJ, Edgar C, Jack CR, et al. Machine learning for comprehensive forecasting of Alzheimer’s Disease progression. Sci Rep 2019;9(1):1–14. PMID:31541187
4. Lee G, Nho K, Kang B, Sohn KA, Kim D, Weiner MW, et al. Predicting Alzheimer’s disease progression using multi-modal deep learning approach. Sci Rep 2019;9(1):1–12. PMID: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 PWN-Polish Scientific Publishers; 2018 Jan 1;38(1):1–15. doi: 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 [Internet] Nature Publishing Group UK; 2020;10(1):1–13. PMID: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 [Internet] BioMed Central Ltd.; 2019 Apr 11 [cited 2020 Dec 23];17(1):1–13. PMID:30971285

8. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna K V, Urban TJ, et al. Machine learning models to predict disease progression among veterans with hepatitis C virus. Nature [Internet] 2009;461(7262):399–401. doi: 10.5281/zenodo.1490242.Funding

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 Elsevier Masson SAS; 2019 Dec 1;103(343):169–179. PMID: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. Circulation [Internet] Lippincott Williams and Wilkins; 1995 Aug 15 [cited 2021 Jan 8];92(4):785–789. doi: 10.1161/01.CIR.92.4.785

11. Olivotto I, Cecchi F, Poggesi C, Yacoub MH. Patterns of disease progression in hypertrophic cardiomyopathy an individualized approach to clinical staging. Circ Hear Fail 2012;5(4):535–546. PMID:22811549

12. O'Mahony C, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi C, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). Eur Heart J [Internet] Oxford University Press; 2014 Aug 2 [cited 2020 Dec 23];35(30):2010–2020. doi: 10.1093/eurheartj/eht439

13. Andy AU, Guntuku SC, Adusumalli S, Asch DA, Groeneveld PW, Ungar LH, Merchant RM. Predicting Cardiovascular Risk Using Social Media Data: Performance Evaluation of Machine-Learning Models. JMIR cardio. 2021 Feb 19;5(1):e24473.

14. Lundberg SM, Allen PG, Lee S-I. A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst. 2017.

15. Magoulas GD, Prentza A. Machine learning in medical applications. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) [Internet] Springer Verlag; 2001 [cited 2020 Dec 23]. p. 300–307. doi: 10.1007/3-540-44673-7_19

16. Wang X, Sontag D, Wang F. Unsupervised learning of disease progression models. Proc ACM SIGKDD Int Conf Knowl Discov Data Min 2014;85–94. doi: 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. Current Cardiovascular

Imaging Reports. 2019 Feb 1;12(2):5.

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. Cardiovascular Interventions. 2019 Jul 22;12(14):1293-303.

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. Proceedings of the National Academy of Sciences. 2021 Jun 15;118(24).

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 18 F-FDG PET of the brain. Radiology Radiological Society of North America Inc.; 2019 Mar;290(3):456–464. PMID:30398430

21. Seah JCY, Tang JSN, Kitchen A, Gaillard F, Dixon AF. Chest Radiographs in Congestive Heart Failure: Visualizing Neural Network Learning. Radiology [Internet] Radiological Society of North America Inc.; 2019 Feb 1 [cited 2020 Dec 23];290(2):514–522. doi: 10.1148/radiol.2018180887

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 [Internet] Radiological Society of North America (RSNA); 2019 May 29 [cited 2020 Dec 23];1(3):e180061. doi: 10.1148/ryai.2019180061

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 [Internet] Springer US; 2019;25(1):70–74. PMID:30617318

24. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc [Internet] American Heart Association Inc.; 2018 Jul 1 [cited 2020 Dec 23];7(13). PMID:29945914

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 [Internet] John Wiley and Sons Ltd; 2020 Jan 12 [cited 2020 Dec 23];22(1):139–147. doi: 10.1002/ejhf.1628

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 Elsevier Ireland Ltd; 2019 Aug 1;288:140–147. PMID: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. Proc - 2017 13th Int Conf Intell Environ IE 2017 Institute of Electrical and Electronics Engineers Inc.; 2017. p. 14–19. doi: 10.1109/IE.2017.19

28. Oguz C, Sen SK, Davis AR, Fu YP, 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 [Internet] BioMed Central Ltd.; 2017 Oct 26 [cited 2020 Dec 23];11(1):1–14. PMID:29073909

29. Rowin EJ, Maron MS. The role of cardiac MRI in the diagnosis and risk stratification of hypertrophic cardiomyopathy. Arrhythmia Electrophysiol Rev [Internet] Radcliffe Cardiology; 2016 Dec 1 [cited 2020 Dec 23];5(3):197–202. doi: 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. W.B. Saunders; 2014. p. 221–230. PMID:24412355

31. Hiemstra YL, Debonnaire P, Bootsma M, Schalij MJ, Bax JJ, Delgado V, Marsan NA. Prevalence and Prognostic Implications of Right Ventricular Dysfunction in Patients With Hypertrophic Cardiomyopathy. Am J Cardiol Elsevier Inc.; 2019 Aug 15;124(4):604–612. PMID:31204037

32. Just E. Understanding Risk Stratification, Comorbidities, and the Future of Healthcare. 2017.

33. Hess OM. Risk stratification in hypertrophic cardiomyopathy. J Am Coll Cardiol [Internet] Elsevier Inc.; 2003 Sep 3 [cited 2020 Dec 23];42(5):880–881. doi: 10.1016/S0735-1097(03)00838-6

34. Steriotis AK, Sharma S. Risk stratification in hypertrophic cardiomyopathy. Eur Cardiol Rev [Internet] Radcliffe Cardiology; 2015 [cited 2020 Dec 23];10(1):31–36. doi: 10.15420/ecr.2015.10.01.31

35. Maron BJ, Rowin EJ, Casey SA, Haas TS, Chan RHM, Udelson JE, et al. Risk stratification and outcome of patients with hypertrophic cardiomyopathy ≥60 years of age. Circulation [Internet] Lippincott Williams & WilkinsHagerstown, MD; 2013 Feb 5 [cited 2020 Dec 23];127(5):585–593. PMID:23275385

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 [Internet] American Medical Association; 2019 Jul 1 [cited 2020 Dec 23];4(7):644–657. PMID:31116360

37. Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, et al. 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 [Internet]. Circulation. Lippincott Williams & WilkinsHagerstown, MD; 2011 [cited 2020 Dec 23]. p. 783–831. PMID:22068434

38. Robnik-Šikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression. Mach Learn Proc Fourteenth Int Conf 1997. p. 296–304.

39. Breiman L. Random forests. Mach Learn Springer; 2001 Oct;45(1):5–32. doi:

10.1023/A:1010933404324

40. Malley JD, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability Machines: Consistent probability estimation using nonparametric learning machines. Methods Inf Med NIH Public Access; 2012;51(1):74–81. PMID:21915433

41. Fabian P, Michel V, Grisel O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. J Mach Learn Res. 2011.

42. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl Discov Data Min New York, NY, USA: Association for Computing Machinery; 2016. p. 785–794. doi: 10.1145/2939672.2939785

43. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. GitHub - keras-team/keras-tuner: Hyperparameter tuning for humans. 2019.

44. Chen S, Bortsova G, Juárez AGU, van Tulder G, de Bruijne M. Multi-task Attention-Based Semi-supervised Learning for Medical Image Segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) Springer; 2019. p. 457–465. doi: 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 Frontiers Media S.A.; 2020 Nov;14:601829. doi: 10.3389/fninf.2020.601829

46. Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G, et al. Semi-supervised learning for network-based cardiac MR image segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) Springer Verlag; 2017. p. 253–260. doi: 10.1007/978-3-319-66185-8_29

47. Viceconti M, Henney A, Morley-Fletcher E. In silico clinical trials: how computer simulation will transform the biomedical industry. Int J Clin Trials Medip Academy; 2016 May;3(2):37. doi: 10.18203/2349-3259.ijct20161408

48. Niederer SA, Lumens J, Trayanova NA. Computational models in cardiology. Nat Rev Cardiol. Nature Publishing Group; 2019. p. 100–111. PMID: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 J Pharmacokinet Pharmacodyn; 2006 Dec;33(6):773–794. PMID:17053984

50. Robnik-Šikonja M. Dataset comparison workflows. Int J Data Sci Inderscience Enterprises Ltd; 2018;3(2):126–145.

51. Bøttcher SG, Dethlefsen C. Deal: A package for learning bayesian networks. J Stat Softw 2003;8:1–19. doi: 10.18637/jss.v008.i20

52. D’Agostino RB, Stephens MA. Goodness-of-fit techniques. New York, NY, USA: Marcel Dekker, Inc.; 1986. ISBN:0824774876

53. Robnik-Šikonja M, Bohanec M. Perturbation-Based Explanations of Prediction Models. Springer, Cham; 2018 [cited 2021 Jan 4]. p. 159–175. doi: 10.1007/978-3-319-90403-0_9

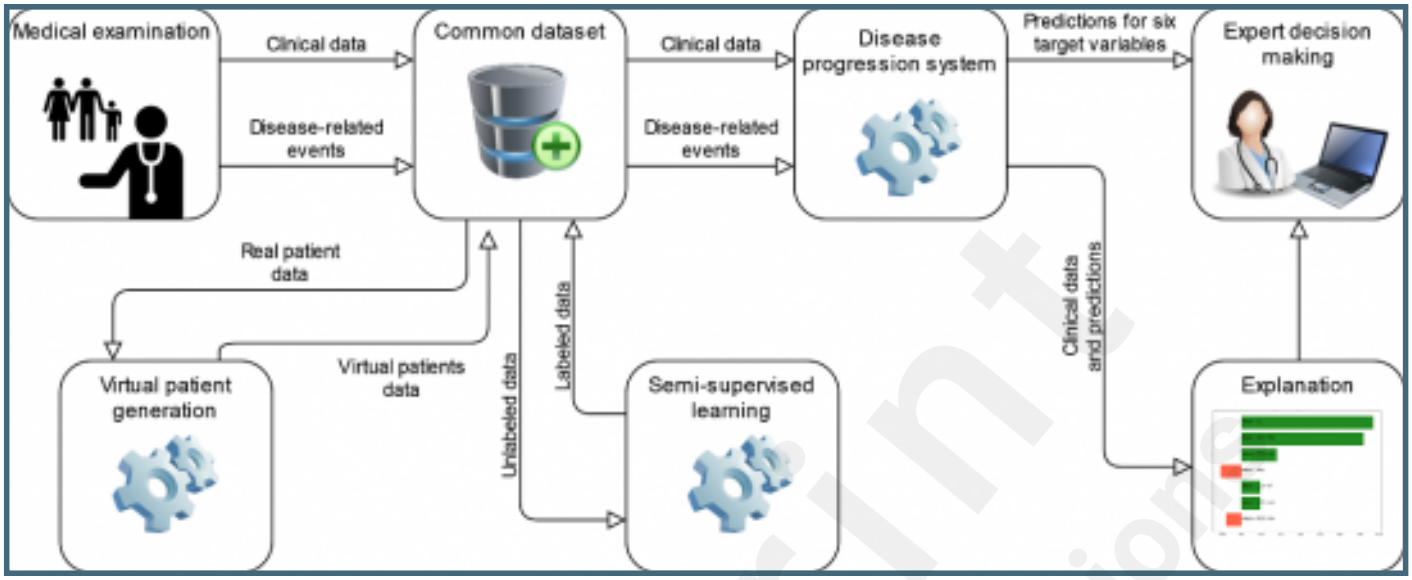
54. SHaRe [Internet]. URL: https://theshareregistry.org/home/patient [accessed 2021-05-14]

Supplementary Files

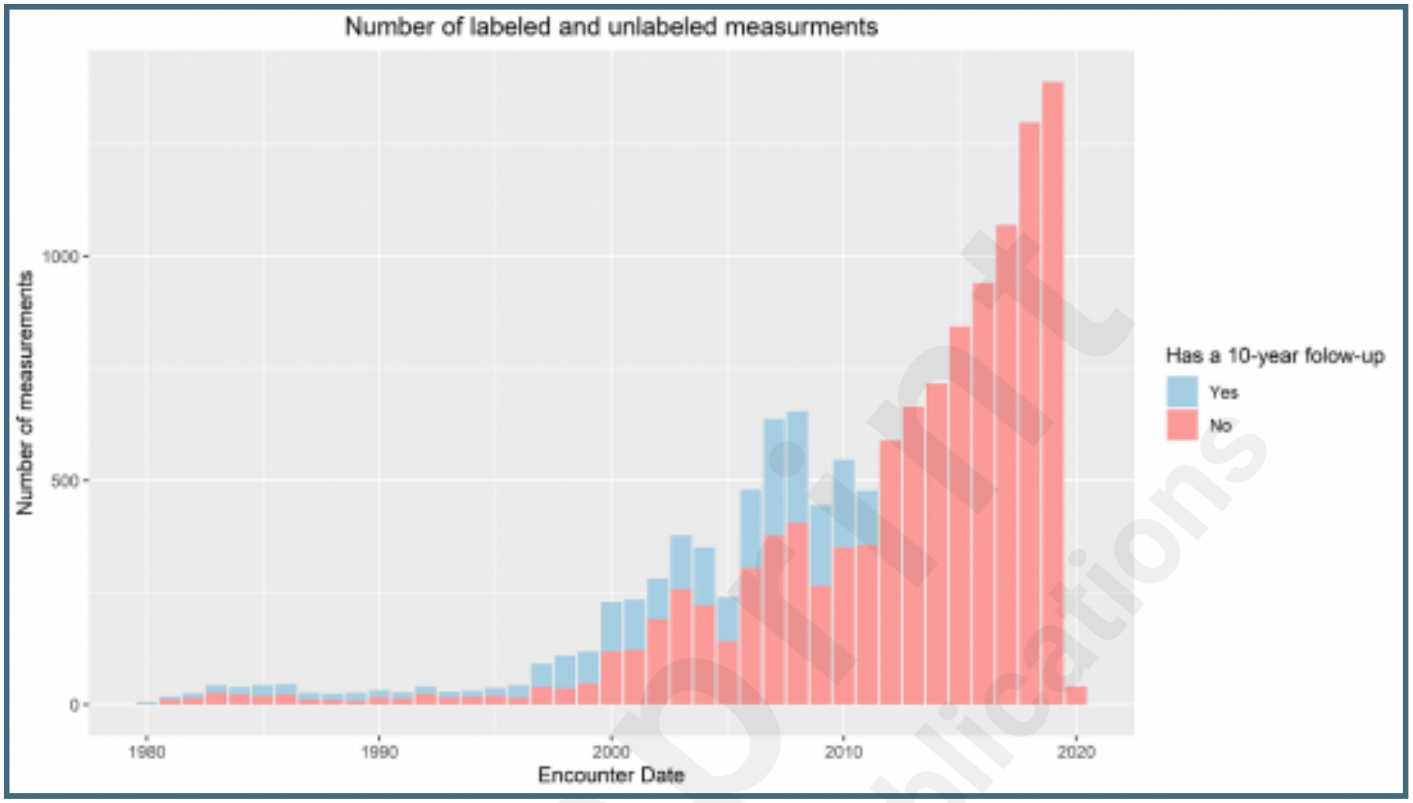
Responses for the reviewers (2nd).
URL: http://asset.jmir.pub/assets/10cadff6d8f5dfc73e3ee266aa38d167.docx

Figures

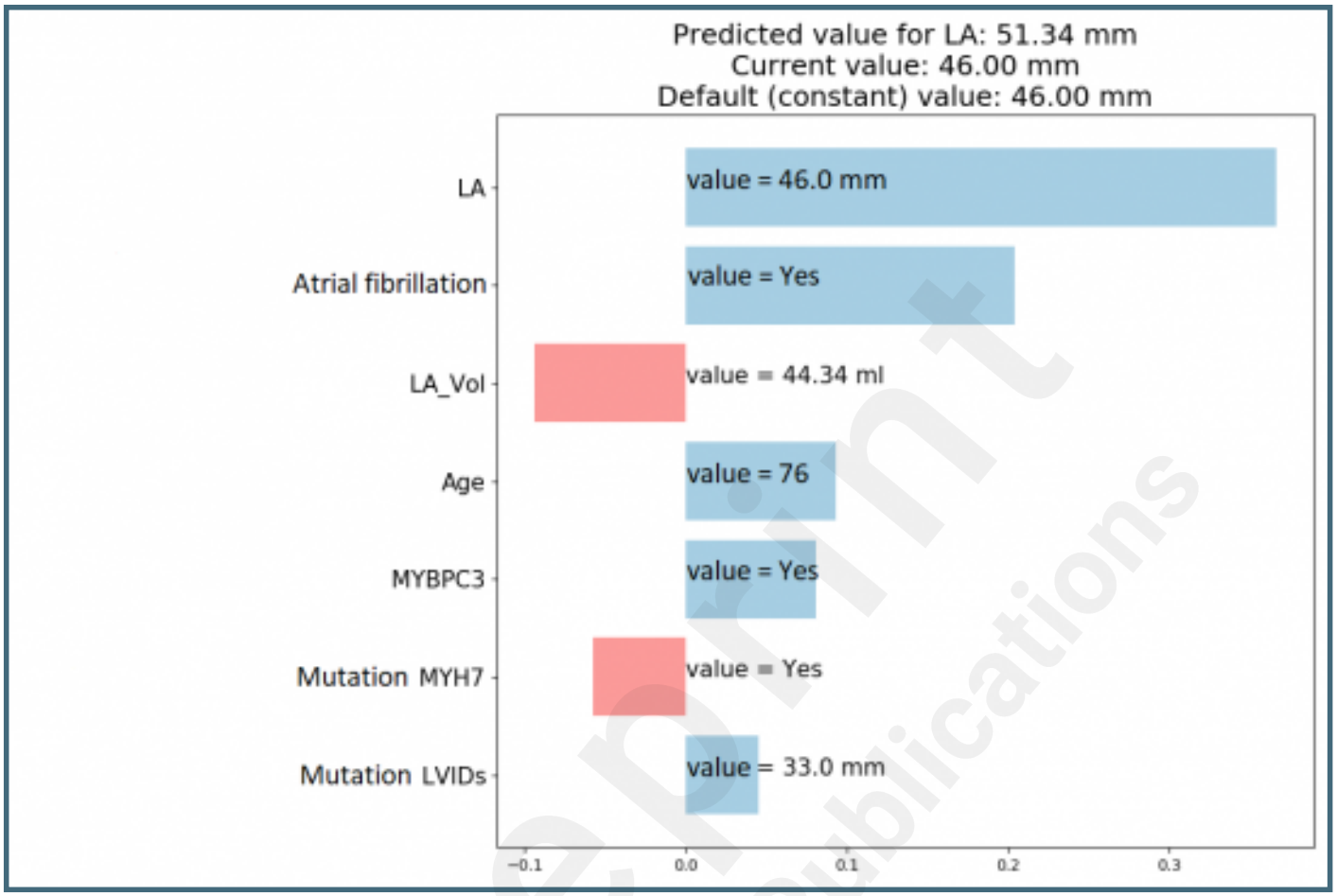
An overview of the proposed disease progression system. The system receives clinical data and disease-related events of a patient as an input, utilizes virtual patient data and semi-supervised learning for self-improvement, and returns the predictions and their explanation for six target variables.



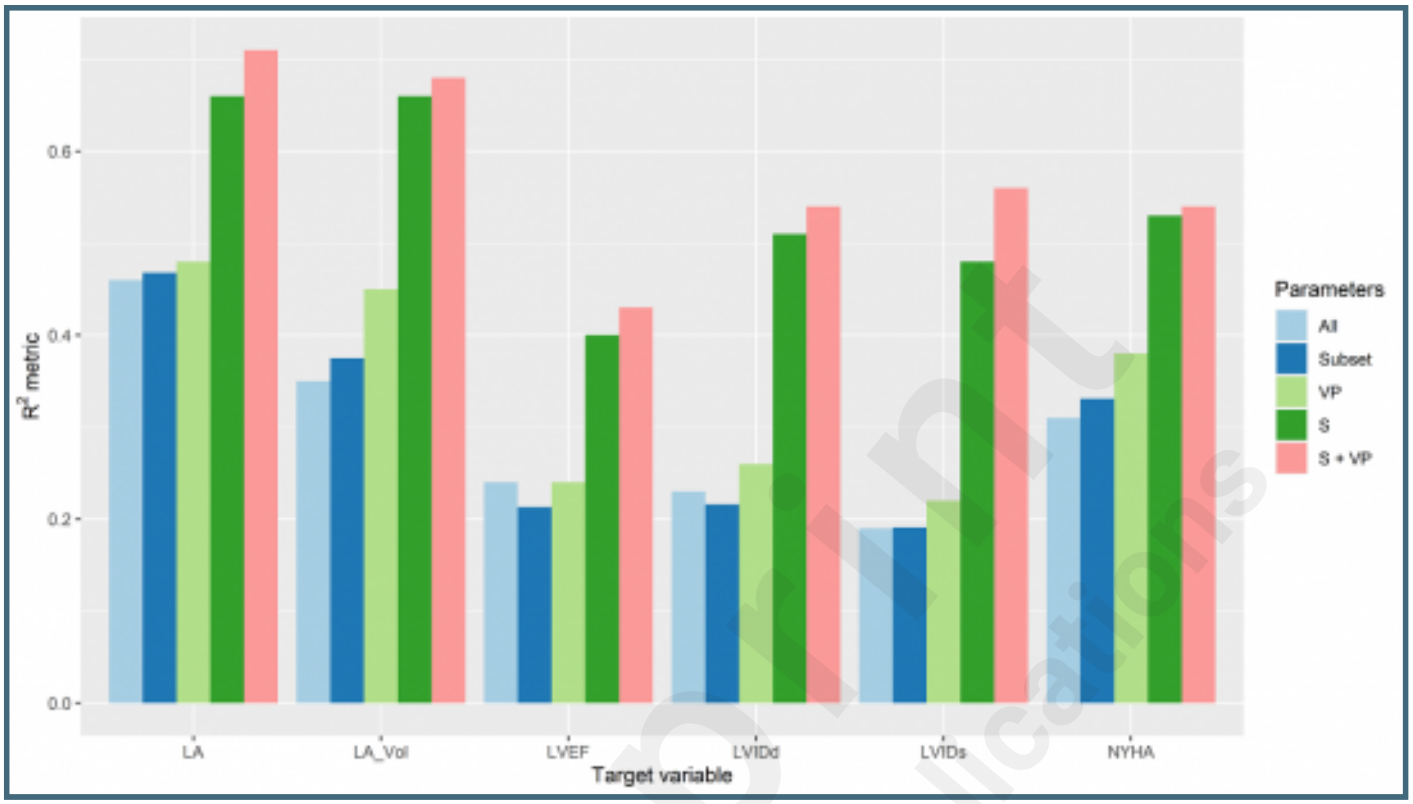
This figure shows relationship between the amount of labeled and unlabeled data. The bars for Yes and No values are stacked, visually revealing the ratio between the labeled and unlabeled data. Note that the rightmost columns do not have 10-year follow up data, as they are younger than 10 years.



An example of explanation of prediction for target variable LA.



The plotted results for R2 statistic for each target variable using different sets (input parameters). Note that VP, S and S + VP are used on feature subsets.



Multimedia Appendixes

Bar graphs of parameter influence for each model used.
URL: http://asset.jmir.pub/assets/88218720a6141553b9ab9fa53b536b57.png

A sample of questionnaire for the first patient.
URL: http://asset.jmir.pub/assets/fc64c156a462edaf01266aeb6464b0bd.pdf

TOC/Feature image for homepages

Placeholder image.