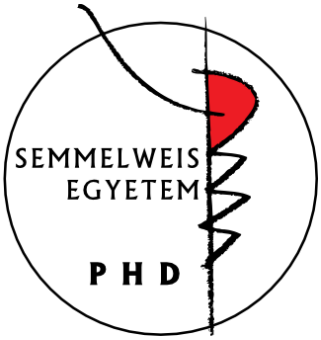
Machine learning-based risk stratification of patients with heart failure

Ph.D. Thesis

Márton Tokodi, M.D.

Doctoral School of Basic and Translational Medicine
Semmelweis University



Supervisor: Attila Kovács, M.D., Ph.D.

Official reviewers: András Horváth, Ph.D.

Pál Kaposi Novák, M.D., Ph.D.

Head of the Complex Examination Committee:

István Karádi, M.D., D.Sc.

Members of the Complex Examination Committee:

Lívia Jánoskuti, M.D., Ph.D.

József Borbola, M.D., Ph.D.

Budapest
2021

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A – late diastolic transmitral flow velocity

ACC – American College of Cardiology

ACE-I – angiotensin-converting enzyme inhibitors

AF – atrial fibrillation

AHA – American Heart Association

AI – artificial intelligence

ANN – artificial neural network

ANOVA – analysis of variance

ARB - angiotensin II receptor blockers

ASIAN-HF – Asian Sudden Cardiac Death in Heart Failure (registry)

AUC – area under the receiver operating characteristic curve

AUPRC – area under the precision-recall curve

BEST – Beta-Blocker Evaluation in Survival Trial

BMI – body mass index

CABG – coronary artery bypass grafting

CAD – coronary artery disease

CCB – calcium channel blockers

CI – confidence interval

CIED – cardiovascular implantable electronic device

CIRF – conditional inference random forest

CKD – chronic kidney disease

CNN – convolution neural network

COMPANION – Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (trial)

COPD – chronic obstructive pulmonary disease

CRT – cardiac resynchronization therapy

CRT-D – cardiac resynchronization therapy-defibrillator

CV – cross-validation

CVA – cerebrovascular accident

DBP – diastolic blood pressure

DD – diastolic dysfunction

DL – deep learning

DM – diabetes mellitus

ECG – electrocardiogram

E – early diastolic transmitral flow velocity

EF – ejection fraction

EHR – electronic health record

e' – early diastolic relaxation velocity

GFR – glomerular filtration rate

GMM – Gaussian mixture models

GWGT-HF – Get With The Guidelines-Heart Failure (risk score)

HF – heart failure

HFpEF – heart failure with preserved ejection fraction

HFrEF – heart failure with reduced ejection fraction

HLD – hyperlipidemia

HR – hazard ratio

HTN – hypertension

IDI – integrated discrimination improvement

INTERMACS – Inter-Agency Registry for Mechanically Assisted Circulatory Support

KNN – k-nearest neighbors

LASSO – least absolute shrinkage and selection operator

LAVi – left atrial volume index

LBBB – left bundle branch block

LV – left ventricular

LVAD – left ventricular assist device

LVEDD – left ventricular end-diastolic diameter

LVEF – left ventricular ejection fraction

LVMi – left ventricular mass index

MACE – major adverse cardiac event

MADIT-CRT – Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy

MAE – mean absolute error

MAGGIC – Meta-Analysis Global Group in Chronic Heart Failure

MASCOT – Management of Atrial Fibrillation Suppression in AF-HF Comorbidity Therapy (trial)

MESA – Multi-Ethnic Study of Atherosclerosis

MI – myocardial infarction

MICE – multiple imputation by chained equations

MIRS – median improvement in risk score

ML – machine learning

MRA – mineralocorticoid receptor antagonists

MSE – mean squared error

NB – naïve Bayes

NEAT-HFpEF – Nitrate's Effect on Activity Tolerance in Heart Failure with Preserved Ejection Fraction (trial)

NRI – net reclassification improvement

NT-proBNP – N-terminal pro-brain natriuretic peptide

NYHA – New York Heart Association

PCI – percutaneous coronary intervention

pEF – preserved ejection fraction

RAFT – Resynchronization-Defibrillation for Ambulatory Heart Failure Trial

RELAX-HF – Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Diastolic Heart Failure (trail)

RMSE – root mean squared error

RNN – recurrent neural network

ROC – receiver operating characteristic

RV – right ventricular

SBP – systolic blood pressure

SEMMELWEIS-CRT – perSonalizEd assessMent of estiMatEd risk of mortaLity With machinE learnIng in patientS undergoing CRT implantation

SHFM – Seattle Heart Failure Model

SMART-AV – SMARTDELAY-determined AV optimization (trial)

SVM – support vector machine

TDA – topological data analysis

TOPCAT –Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (trial)

TRF – traditional random forest

TRV – tricuspid regurgitation peak velocity

1. INTRODUCTION

1.1 The advent of machine learning in cardiovascular medicine

With data-rich technologies and data-heavy areas of biomedical science entering the clinical arena, physicians are being inundated with a staggering volume of data requiring more sophisticated interpretation while being expected to perform more efficiently. In other words, the ever-growing complexity of medicine now exceeds the capacity of the human mind (1). A potential solution is machine learning (ML) – a subfield of artificial intelligence (AI) – that can support clinical decision-making to enhance every stage of patient care, from research to diagnosis and tailoring therapeutic strategies (2, 3).

Originated from the vision of Alan Turing and Marvin Minsky that a machine could imitate human intelligence (4, 5), the field of AI research was officially born in 1956 at the Dartmouth workshop, where it gained its vision, mission, and hype (6, 7). Since then, AI has endured a bumpy journey with several hype cycles and survived two major droughts of funding and interest in 1974 – 1980 and 1987 – 1993 (often referred to as "AI winters"). Emerging from AI, the field of ML was coined by Arthur Samuel in the 1950s and started to flourish in the 1990s when its focus was shifted from the symbolic and less-specified approaches inherited from AI toward methods and models adopted from statistics and probability theory to address well-defined tasks (8, 9). Currently, AI and ML are attracting increased attention which is attributable to three key factors: (1) the increasing ubiquity of large, multifaceted data sets, (2) the availability of relatively inexpensive and powerful computational resources, and (3) advances in algorithms that allow for efficient training of deep neural networks.

ML has been woven into the fabric of our everyday life and has sparked tremendous innovation in the realms of business, entertainment, and technology. Although its application to medicine has been less apparent for many years, the ML community has recently been concentrating its efforts on the healthcare sector, and a plethora of studies has been published over the past decade exploring the potential utility of ML in various clinical scenarios (10). As ML-based algorithms permeate into clinical cardiology, they will facilitate the scheduling and protocolling of medical tests and appointments, help remote monitoring of patients through wearable devices, enhance the interpretation of electrocardiograms (ECGs), enable automated measurements from

echocardiograms, computed tomography, and cardiac magnetic resonance scans, perform risk stratification using every piece of data available in the patient's chart, and guide therapeutic decisions (10-15). Designed, validated, and implemented appropriately, ML models will help us acquire, interpret, and synthesize healthcare data from diverse sources and place it at our fingertips (10). Thus, ML holds promise to revolutionize medical research and clinical care leading to an optimized day-to-day clinical workflow with improved diagnostics, risk assessment, and ultimately outcomes.

1.2 Definition of artificial intelligence and machine learning

AI is a broad and ambiguous term that describes any computational system simulating and mimicking human intelligence. AI can indicate general-purpose AI (often referred to as artificial general intelligence), in which the system is self-sufficient and possesses cognition comparable to that of humans. Yet, such general AI has not been invented, and only the so-called narrow AI – designed for solving single, specific tasks (e.g., playing a game or driving a car) – is available currently.

ML, traditionally considered a branch of AI, is an emerging technology paradigm that enables computers to learn sophisticated patterns and insights from the data without being explicitly programmed. Thus, ML allows the user to glean knowledge from pre-existing data and apply it to future predictions. It is essential to highlight that ML does not constitute general intelligence as it is designated to tackle well-defined problems that are considered overly difficult to solve using rule-based algorithms. Unlike the traditional programming paradigm where computer programs perform tasks according to a pre-defined set of rules created based on human experience and knowledge, ML aims to generate a model from the input (and in some cases, the desired output) data to perform a given task.

1.3 Machine learning vs. conventional statistics

Predictive models can be constructed using statistical methods or ML techniques (16, 17). Although these two approaches have significant areas of overlap, there are fundamental differences in their motivating philosophies. Traditional statistical models aim to obtain causal inferences about a population given a data sample and a set of assumptions, whereas ML focuses on algorithmically representing data structure and

providing more accurate predictions. Nonetheless, as these two ambitions are often intertwined, it is impossible to place a definite boundary between conventional statistics and ML methods. Instead, we should view them as analogous that are often applied to answer different questions. Although statistical modeling does not strictly focus on making predictions, it can be used to do so. For instance, logistic regression and linear regression are statistical models that are commonly used in ML for predictive purposes, and they frequently serve as reference models for the comparison of more sophisticated ML algorithms.

While developing predictive models, there are three major analytic challenges that we have to overcome (16): (1) non-linearity – the predictor's effect on the outcome does not change uniformly throughout its range (e.g., age vs. mortality, body mass index [BMI] vs. occurrence of several diseases), (2) heterogeneity of effects – the predictor's relationship with the outcome depends on the level of some other variables (e.g., gene-environmental interactions), and (3) the vast amount of potential, often correlated predictors. Not accounting properly for these factors may result in significantly lower prediction accuracy. If our goal is to generate a model that provides highly accurate predictions, ML algorithms can be advantageous over conventional statistical models as they were designed to confront the aforementioned issues at the sacrifice of the interpretability of the relationship between risk factors and the outcome of interest.

1.4 The machine learning analysis pipeline

The ML analysis pipeline comprises discrete steps (Figure 1): (1) data collections, (2) data pre-processing, (3) choosing an ML algorithm, and (4) model training, validation, and testing.

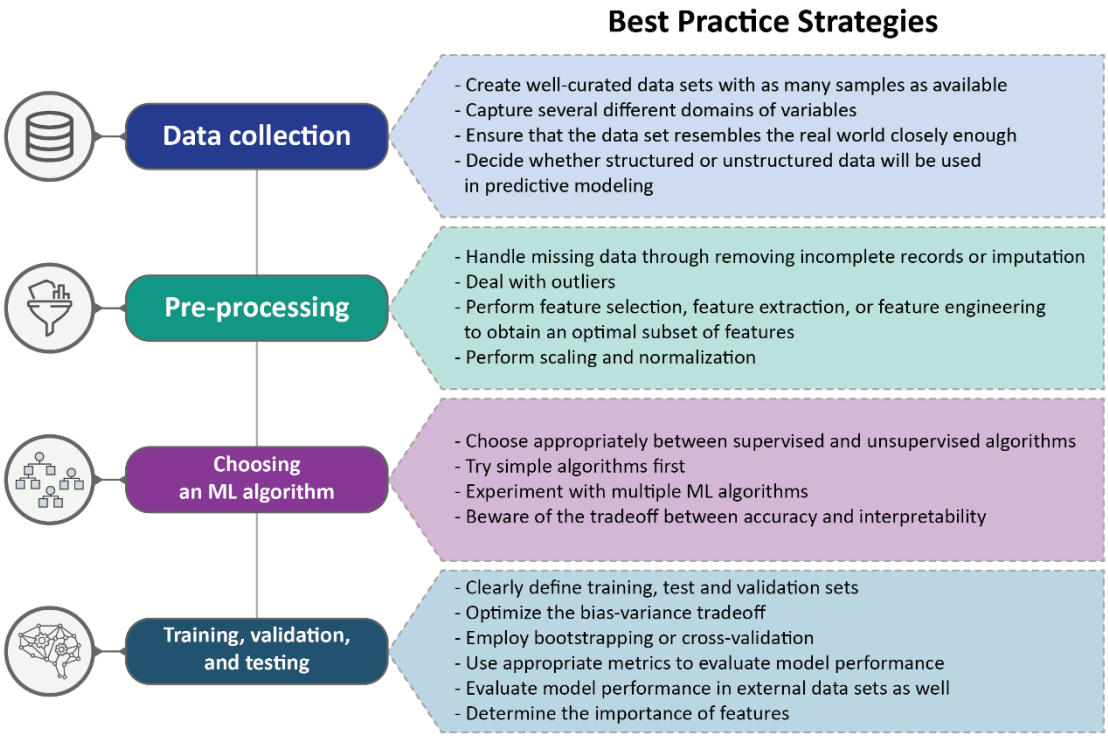


Figure 1 *The machine learning analysis pipeline*

The ML analysis pipeline comprises four consecutive steps: (1) data collection, (2) data pre-processing, (3) choosing a machine learning algorithm, and (4) model training, validation, and testing.

ML – machine learning

1.4.1 Data collection and database building

The power of predictive models heavily depends on having well-curated, sufficiently large data sets. ML-based modeling can be performed with only a few hundred patients; however, more extensive data sets would significantly enhance the predictive power and generalizability. The quality, accuracy, and richness of input features will also impact the effectiveness of the predictive models. Provision of inappropriate, insufficient, or incorrectly categorized data leads to a data set that does not resemble the real world closely enough for ML to create a representative model, resulting in poor prediction performance. Researchers should endeavor to capture several different domains of variables (e.g., clinical, imaging, and genomic covariates) as holistic models incorporating several different features tend to be more robust. When possible, increasing the number of data samples often decreases prediction error and improves performance, but gathering more data is not always feasible or cost-effective in medical research.

Input data can be structured or unstructured. Currently, the majority of medical research is done using structured data, which are organized into structured frameworks with rows corresponding to subjects (i.e., patients) and columns representing features. On the other hand, unstructured data are stored without a well-organized structure. In medical research, textual information in electronic health records (EHRs), radiological images, audio and visual clips are examples of data that are considered unstructured. The emergence of ML (especially deep neural networks) has opened avenues to effectively analyze such data, which are thought to contain up to 90% of all potentially usable information and to be a massive resource for medical research (18).

1.4.2 Data pre-processing

Given the number of imperfections that can contaminate medical data sets such as missing values, sparsity, presence of outliers, and inter-variable differences in scale, pre-processing steps – e.g., imputation of missing data and normalization – are necessary prerequisites for ML algorithms.

Clinical data sets often contain missing values; thus, clinicians and data scientists must decide how to handle them (e.g., by removing incomplete records or imputing values for missing data) while avoiding the excessive loss of data and maximizing both applicability and accuracy model. The mechanism of missingness is essential to be established in order to determine whether the missingness is ignorable (it is not related to the unobserved value, i.e., missing at random) or non-ignorable (it is related to the unobserved values). It is always challenging to deal with non-ignorable missing data; however, if missingness is ignorable, it can be addressed by imputation. Several imputation methods exist, ranging from single-value imputation to more sophisticated techniques, such as multiple imputation by chained equations (MICE) (19, 20).

In many cases, data collection will yield a vast amount of variables, which often exceeds the sample size of the cohort used for modeling. Utilizing all features would inevitably lead to overfitting (i.e., the mismatch between the complexity of the ML algorithm and the size of the provided training data set); therefore, an optimal subset of input variables is required to be identified. Numerous strategies are available that can be applied to overcome the curse of dimensionality, comprising the full gamut from feature selection (pinpointing features that are robust, informative, and non-redundant using

filtering, wrapper, and embedded methods) and feature extraction (projecting multi-dimensional feature vectors to a lower dimensionality subspace allowing for a more concise representation of each object) to feature engineering (handcrafting new features with unsupervised learning techniques).

If the range of different continuous features varies widely, scaling transformations can be applied to eliminate the possibility of model bias caused by the differing magnitude of the numerical values. Scaling is a crucial pre-processing step for ML algorithms like support vector machines (SVMs) and k-nearest neighbors (KNN) classifiers, where the distance between the data points is of particular importance. The two most frequently employed scaling methods are standardization and min-max scaling. Standardization (often referred to as Z-score transformation) is a technique that transforms the data so that the mean of the observed values is 0 and the standard deviation is 1 (Equation 1).

$$z = \frac{x - \mu}{\sigma}$$

Equation 1 *Standardization (Z-score transformation)*
x – the original value of the given observation, z – the standardized (z-transformed) value of the given observation, μ – the mean of the sample, σ – the standard deviation of the sample

In min-max scaling, values are shifted and rescaled so that they end up ranging between 0 and 1 (Equation 2). The major drawback of bounding the data to such a small fixed range is that the weight of outliers will be suppressed.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Equation 2 *Min-max scaling*
x – the original value of the given observation, x' – the min-max scaled value of the given observation, x_{min} – the minimum of the sample, x_{max} – the standard deviation of the sample

In many modeling scenarios, normal distribution of input features is desirable. A wide variety of transformers exist that can be utilized to map data from any distribution to as close to a normal distribution as possible in order to stabilize variance and minimize skewness. The most commonly used ones include power transforms such as Box-Cox, Yeo-Johnson, log, and quantile transforms. Nevertheless, any function can be converted into a transformer to assist in this step of pre-processing.

1.4.3 Choosing a machine learning algorithm

Once the data set is finalized and pre-processed, we can proceed to choose an ML algorithm appropriate for the type of problem that we are trying to solve. Different ML algorithms are associated with distinct strengths and limitations; therefore, the choice of modeling has a definite impact on prediction performance. Multiple-modeling implementations are desirable; therefore, a series of experiments with various ML algorithms should be performed to determine which algorithm has the highest predictive power in the given scenario (16).

In ML, three basic learning paradigms exist: (1) supervised, (2) unsupervised, and (3) reinforcement learning. Within each of these categories, there are numerous ML algorithms with different underlying structures aiming to address particular research questions based on different data complexities. Of note, the aforementioned approaches to ML do not encompass all techniques, and there are methods that do not fit neatly within these paradigms, such as active learning and one-shot learning. However, these techniques are less frequently used in cardiovascular research.

While choosing an optimal ML algorithm, we should be aware of the tradeoff between accuracy and interpretability. Thus, the question arises: should we use a black-box algorithm with high accuracy or an easily interpretable algorithm with only modest accuracy? The lack of interpretability of ML-based decision-making implies that it can be challenging to verify whether the learned rules have truly generalized to real-life clinical situations. Conceivable solutions include the substitution of black-box algorithms with interpretable models or the utilization of model-agnostic methods to explain the decision of the black-box with case- and model-specific explanations (21).

Although the detailed description of each ML algorithm goes beyond the scope of this section, the concepts of the most frequently applied techniques will be briefly discussed.

1.4.3.1 Supervised learning

Supervised ML algorithms attempt to model how independent variables relate to a dependent variable (i.e., the label of interest). The reference to supervision indicates whether data have been labeled with the true response or outcome (the ground truth), and algorithms that use labeled data are termed supervised learning techniques – a collection

of ML approaches in which the algorithm learns directly from large quantities of correctly labeled examples.

These algorithms can be applied to tackle both classification and regression tasks. In the former case, the goal is to correctly assign a binary or multi-class label, while in the latter, it is to accurately predict a real-valued output. Although many algorithms are flexible enough to accomplish both types of analysis with only slight modifications, constraints such as interpretability, computational cost, and the type of available data need to be considered in tailoring the choice of algorithm.

Besides categorical and continuous labels, many algorithms are applicable to time-to-event outcomes as well. Such approaches are particularly useful when there is loss to follow-up and censoring. Moreover, such models allow researchers to derive risk prediction over multiple time points. Nevertheless, some studies suggest that when we aim to predict the risk at a specific time point, binary models perform better, likely due to the non-proportionality in predictor effects (22). This is particularly true when complete follow-up data is available for all individuals.

1.4.3.1.1 Regularized regression

If a linear model contains a vast amount of predictor variables or if these predictors are strongly correlated, the standard ordinary least squares parameter estimates have large variance. Thus, the model is considered unreliable. To tackle these challenges, several solutions have been proposed that directly manipulate traditional regression models. The simplest amendment to regression is (forward or backward) stepwise feature selection. However, as the final fit is based on a regression model, the limitations of regression persist, and these models are still prone to overfitting (i.e., the model learns the details and noise in the training data to the extent that it hampers the performance of the model on new data).

An extension of selection methods, instead of choosing the optimal subset of variables, is to regress an outcome onto all of the predictors. To handle many predictor variables, these methods shrink the regression coefficients toward zero. This shrinkage is achieved by placing a penalty on the summation of the estimated coefficients. Although this shrinkage results in biased regression, it ultimately leads to a more stable model that produces better predictions, particularly when applied to external data sets. In other

words, variance is reduced at the cost of introducing some bias. This approach is called regularization and is almost always beneficial for the predictive performance of the model.

Regularized regression methods differ in the way they perform the aforementioned shrinkage. In ridge regression, the loss function is augmented in such a way that not only is the sum of squared residuals minimized, but also the size of parameter estimates is penalized in order to shrink them toward zero (23, 24). The least absolute shrinkage and selection operator (LASSO) regression is relatively similar conceptually to ridge regression (25). It also adds a penalty for non-zero coefficients, but unlike ridge regression which penalizes the sum of squared coefficients (L2 penalty), LASSO penalizes the sum of their absolute values (L1 penalty). Thus, coefficients can be set exactly to zero in LASSO, which is never the case in ridge regression. Simply put, LASSO regression results in a full shrinkage of a subset of variables, effectively performing feature selection.

LASSO tends to do well if there are only a few significant variables among the input features (i.e., only a few predictors actually influence the response). On the other hand, ridge regression works well if most predictors impact the response substantially. To amalgamate the favorable characteristics of both ridge and LASSO regression, elastic-net regression was introduced that linearly combines the penalties of these two methods (26). Therefore, it performs feature selection and regularization simultaneously, and it is the appropriate regression method of choice if myriads of features are included that form groups of highly correlated variables.

1.4.3.1.2 Tree-based algorithms

Tree-based methods represent a widely applied class of powerful but deceptively simple algorithms. A decision tree is a flowchart-like structure that can be used to predict a categorical (classification tree) or numerical outcome (regression tree) (27, 28). It comprises decision nodes, branches, and leaf nodes (Figure 2A). The most superiorly located decision node is usually referred to as the root node. Each decision node represents a decision criterion and has branches corresponding to the outcomes of that decision. The leaf (or terminal) nodes are the final nodes that carry the predicted value or class and do not contain any further branches. Classification and regression trees mimic

how a physician may approach a patient with a series of subsequent questions based on the answers to the preceding ones.

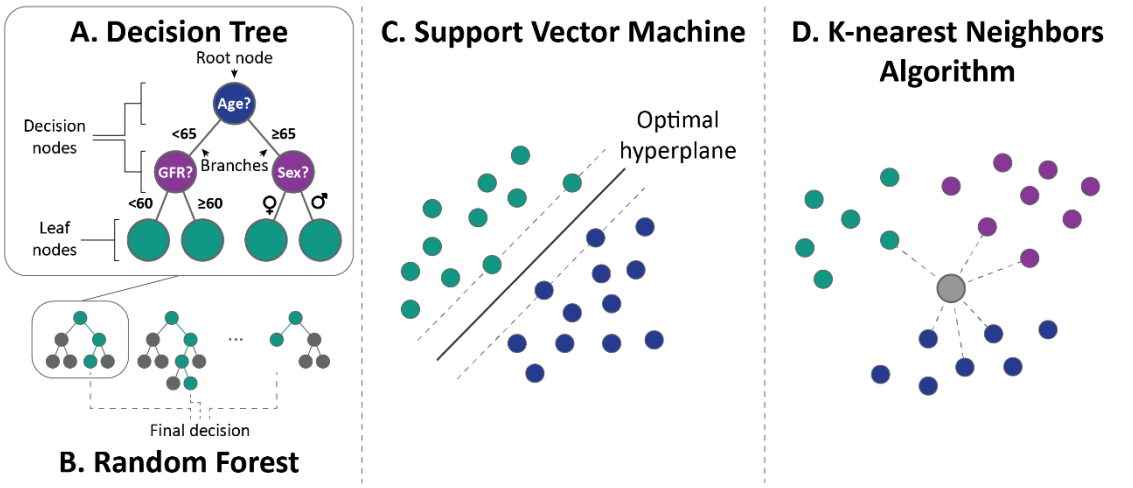


Figure 2 *Examples of supervised machine learning algorithms*

(A) An example of a decision tree, (B) a schematic illustration of a random forest, (C) a schematic illustration of a support vector machine, and (D) a schematic illustration of the k-nearest neighbors algorithm, $k = 6$.

GFR – glomerular filtration rate

To build a classification or regression tree, a search is performed recursively among the available predictors to identify the one which best separates the data points (i.e., patients) into two groups with the most disparate outcome of interest. Since each split is binary, non-linearity can be captured in the data, as multiple splits on the same predictor can occur within the same tree. The iterative searching for the optimal predictor at each split continues until all groups (i.e., nodes) are homogenous. To create a more stable tree, the algorithm then prunes the tree to reduce the complexity and over-specification of the model.

Although classification and regression trees are able to handle non-linearity, interactions, and an immense number of predictors, they are high-variance learners and do not generalize well to external data sets, limiting their utility as stand-alone prediction models. Nonetheless, it is possible to improve the overall prediction performance by aggregating the results from multiple trees through so-called ensemble methods. One of the most common ensemble methods with trees is random forest (Figure 2B), which uses the bagging procedure to combine multiple trees (29, 30). The random forest approach

also uses another trick to decrease the correlation between the fitted trees: when growing each tree, instead of only sampling over the observations in the data set to generate a bootstrap sample, sampling is also performed over features and only a randomly selected subset of them is kept to build a tree. A major advantage of sampling over the feature space is that it makes the model more robust to missing data as observations with missing data can still be regressed or classified based on the trees that take into account only features where data are available.

Besides random forest, boosting-based methods (e.g., adaptive boosting and gradient boosting) represent another subset of ensemble methods combining multiple decision trees (31, 32). Boosting is the concept of fitting models iteratively such that the model training at a given step depends on the models fitted at the previous steps (31). Unlike bagging-based methods in which weak learners are fitted in parallel (independently from each other), boosting-based methods fit weak learners in a sequential manner.

Ensemble methods can be conceptualized as forming a robust overall prediction by aggregating the predictions of many simpler and weaker predictive models such as decision trees. At testing, predicted results are obtained from each decision tree. Then, the class receiving the majority of votes made by each decision tree is selected as the final predicted class. In the case of a regression task, where the output class is numerical, the mean value of the predicted outputs is used. This is similar to the process of deriving a clinical diagnosis for a complex patient by utilizing consultations from many specialists, each of which would look at the patient in a slightly different way. Thus, their combined decision would often be better than a single physician's decision alone.

1.4.3.1.3 Support vector machines

SVMs are widely used algorithms in the cardiovascular domain. These algorithms project observations into a higher dimensional space via mappings known as kernels, and then they attempt to identify a boundary (i.e., hyperplane) that maximizes the separation between the classes (Figure 2C) (33). The term "support vectors" refers to data points that are located the closest to the hyperplane and significantly influence the position and orientation of the hyperplane. Therefore, they are crucial elements of the data set: these are the data points that facilitate the maximization of the classifier's margin. The SVM

kernel is a function that aims to convert a not separable problem to a separable problem by taking a low dimensional input space and transforming it to a higher-dimensional space. The most commonly used kernel functions include linear, polynomial, and Gaussian radial basis function kernels. The last two are mainly useful to solve non-linear separation problems.

Despite their popularity and their ability to capture non-linear relationships, SVMs have two major downsides. First, they perform non-probabilistic classification; thus, they are better suited for categorical outcomes. Nonetheless, there are secondary methods (e.g., Platt scaling or isotonic regression) that can circumvent this limitation by computing probabilistic outcomes. Second, the computation of the input observations in a very high-dimensional space can be difficult or even impossible.

1.4.3.1.4 K-nearest neighbors algorithm

The KNN algorithm is a simple, easy-to-implement supervised ML algorithm that can be used to solve both classification and regression tasks (34). It assumes that similar objects exist in close proximity. It compares each data point to others via a distance function to identify the k nearest (i.e., most similar) data points (Figure 2D). Then, a label is assigned by majority voting (in case of classification) or by averaging the labels of the identified neighbors (in case of regression). In the medical context, KNN can be thought of as predicting the patient's outcome based on previous patients experiencing similar symptoms. The most commonly used distance metric is Euclidean distance; however, there are many others, such as the Manhattan or Minkowski distance.

During model training, emphasis should be placed on the optimization of k. If k is too small, the model will be more sensitive to outliers. Inversely, with the increasing value of k, predictions become more stable and accurate due to majority voting / averaging. However, if k is too large, the neighborhood may include too many less relevant data points leading to decreased predictive power.

KNN's main drawback is that it becomes significantly slower as the volume of data increases. Thus, it is an impractical choice in environments where predictions need to be made rapidly.

1.4.3.1.5 Naïve Bayes classifiers

Naïve Bayes (NB) classifiers are a set of supervised learning algorithms based on the Bayes' theorem, which describes the probability of an event based on prior knowledge of conditions associated with that particular event (35, 36). These algorithms assume independence between every pair of input features. This is a very strong (i.e., naïve) assumption that is most unlikely in real-world data. Nevertheless, NB classifiers may perform surprisingly well on data where this assumption is not met. Although the NB learning scheme is primarily designed for classification tasks, there are solutions that enable its application for regression tasks as well (37).

There are different types of NB classifiers, each suited for input features with specified distribution, such as Gaussian NB (for real-valued features distributed according to Gaussian distribution), multinomial NB (for data exhibiting multinomial distribution), or Bernoulli NB (for data distributed according to multivariate Bernoulli distribution). Bernoulli and multinomial NB are typically employed for document classification, where the input features are the presence of a term (in Bernoulli NB) or the probability of a term (in multinomial NB) (38).

1.4.3.2 Unsupervised learning

As opposed to supervised algorithms, unsupervised learning methods use unlabeled data, meaning there are no outcomes or prediction labels assigned to the data points. Thus, instead of fitting data to a pre-specified outcome, these algorithms attempt to identify any potentially consistent, underlying patterns in the data.

Common applications of unsupervised algorithms include (1) clustering analysis, where data are grouped according to similar characteristics; (2) density estimation, where data are analyzed to estimate its probability distribution and to perform anomaly detection; and (3) dimensionality reduction, where the number of input variables is reduced to a set of core features that capture the essence of the data while reducing linearity or redundancy between features. The last of these applications, dimensionality reduction, plays an important role in data visualization and circumventing the curse of dimensionality. In the following section, the most frequently used clustering algorithms will be briefly reviewed.

1.4.3.2.1 Clustering algorithms

Cluster analysis is an unsupervised ML technique utilized to create homogenous groups (i.e., clusters) based on hidden patterns in the data without a priori knowledge. It can be used as a stand-alone tool for exploration or hypothesis generation. Nevertheless, clustering often represents only the first step to group the data before a more insightful analysis is performed. The identified clusters can be used to label data points; then, the newly labeled data can be exploited by a subsequent supervised learning step.

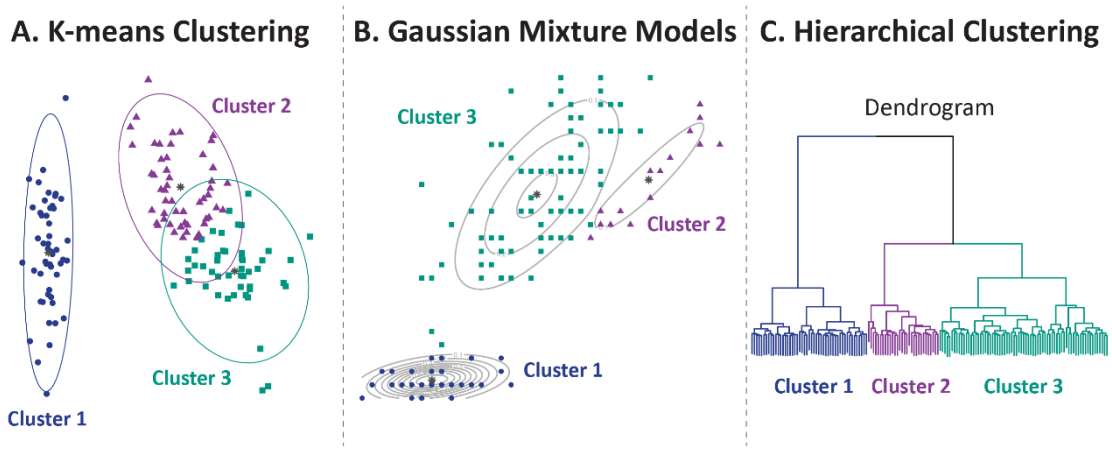


Figure 3 *Examples of clustering algorithms*

(A) An example of k-means clustering, (B) an example of Gaussian mixture models, and (C) a dendrogram visualizing the process and the result of (agglomerative) hierarchical clustering.

The figure was created in R (version 3.6.2, R Foundation for Statistical Computing, Vienna, Austria) using the iris data set.

Clustering algorithms may enable precision cardiology by synthesizing multiple sources of evidence to refine monolithic disease categories (e.g., heart failure [HF]) into more stratified and ultimately more personalized disease concepts. This idea is often termed precision phenotyping. The generated clusters may lead to novel groups of data points that may illuminate novel subtypes of a disease, new biomarkers, or new predictors of a clinical outcome of interest. Importantly, clusters that may suggest new groupings of disease or other novel insights must be validated clinically, as it can be easy to adjust clustering parameters to create subgroups or lump together groups that should be separated.

K-means clustering is one of the simplest and most extensively used unsupervised ML algorithms (Figure 3A). It starts with a pre-defined number (k) of randomly selected

centroids and then performs iterative calculations to minimize the sum of distances between the data points and their respective cluster centroid (39). The algorithm halts optimizing clusters when either the centroids have stabilized or the pre-specified number of iterations has been achieved. The centroids of the k clusters can be used to label new data points. Despite its popularity, k-means clustering has several important limitations that should be acknowledged. First, it assumes that the clusters are spherical and evenly sized. Therefore, if the natural clusters occurring in the data set are non-spherical, k-means clustering is probably not a good choice. Second, k-means clustering requires prior knowledge of k (i.e., the number of clusters). Finally, its results are highly dependent on the initial position of the centroids. Thus, the results may not be reproducible and may lack consistency.

Gaussian mixture models (GMMs) can also be used to cluster unlabeled data (Figure 3B). GMMs are probabilistic models assuming that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters (40). They can be thought of as generalizing k-means clustering to incorporate information about the covariance structure of the data and the centers of the latent Gaussians. Due to this fact, GMMs are expected to mimic the data better than k-means clustering. Moreover, as GMMs contain a probabilistic model under the hood, they are also capable of providing probabilistic cluster assignments, whereas k-means clustering performs a hard classification (i.e., a data point is deterministically assigned to only one cluster).

Hierarchical clustering algorithms represent another frequently used class of clustering methods (Figure 3C). This clustering technique has two subtypes: (1) agglomerative hierarchical clustering and (2) divisive hierarchical clustering. Agglomerative hierarchical clustering is a bottom-up approach: each observation starts in a single-point cluster, then pairs of the closest clusters are recursively merged until there is only one cluster left. There are several methods to decide which clusters to merge, such as the complete linkage method (computes all pairwise dissimilarities between the elements of two clusters and considers the largest value of these dissimilarities as the distance between the two clusters) or Ward's minimum variance method (minimizes the total within-cluster variance). In contrast to agglomerative hierarchical clustering, divisive hierarchical clustering works in a top-down manner. It begins with a single

cluster that contains all data points. At each iteration step, the most heterogeneous cluster is divided into two. The partitioning is performed recursively until all data points are in their own cluster. The process and results of hierarchical clustering are usually visualized via dendrogram – a type of tree diagram showing hierarchical relationships within the data set (Figure 3C). Importantly, unlike centroid-based clustering algorithms, hierarchical clustering cannot be easily used to classify new data points. A solution is to create a classifier on top of the labeled data given by hierarchical clustering.

Besides the aforementioned techniques, there are many other clustering algorithms, such as density-based spatial clustering of applications with noise or mean shift clustering.

1.4.3.3 Semi-supervised learning

In the previous two learning paradigms, labels were provided for all observations (supervised learning), or no labels were available at all (unsupervised learning). Semi-supervised algorithms fall in between these two approaches. If labels are absent for the majority of observations, semi-supervised algorithms are the methods of choice for model creation. These methods exploit the idea that despite the unknown group membership of unlabeled data, such observations still bear relevant information about group parameters.

1.4.3.4 Reinforcement learning

Reinforcement learning algorithms use iterative information about the outcome of their predictions in a computationally simulated environment as feedback to improve their future predictions (i.e., to maximize the reward or minimize the risk). In the process, the algorithm learns from its experience until it explores the full range of possible states. Reinforcement learning allows machines and software agents to automatically determine the ideal behavior within a specific context to maximize its performance.

Reinforcement learning has been successfully applied in robotics, gaming, or self-driving cars (41-46). Moreover, its potentials have been explored in some clinical tasks, such as guiding dofetilide dosing or ventilator settings (47, 48). However, reinforcement learning may be of limited use in clinical tasks where the goal and the environment's possible responses are much more complex.

1.4.3.5 Neural networks and deep learning

Artificial neural networks (ANN) are computational models consisting of several processing elements (i.e., neurons) that receive inputs and deliver outputs based on predefined activation functions. They consist of an input layer, one or more hidden layer(s), and an output layer, where input and output layers indicate the original data and the output of the algorithm, respectively. ANN is often referred to as feed-forward neural network as it propagates the weighted and combined activation signals from one layer of nodes to the next in one direction without cyclical connections. Based on the number of hidden layers, ANNs can be categorized as shallow (1 hidden layer) or deep (>1 hidden layer). Deep learning (DL) – emerged in the past decade, in part because of the advent of graphics processing unit-based parallel computing – refers to having more than one hidden layer between the inputs and outputs of the algorithm, which allows for greater abstraction at the cost of speed and computational power. Through multiple hidden layers, the raw input is gradually converted into more abstract and essential features that represent the original data. As such, DL extracts key features from raw data and returns outputs as classification or regression. In many instances, DL techniques offer a powerful alternative to conventional ML, enabling the user to perform more complex analyses.

Two other common forms of DL models for supervised learning are convolutional neural networks (CNN) and recurrent neural networks (RNN). CNNs are a class of neural networks aptly suited for image analysis. These algorithms require minimal pre-processing and are generally composed of multiple convolutional layers, where multiple filters of shared weights are used to find local patterns in organized data such as images. RNNs are designed for processing temporal information or sequential data such as speech and language. They reuse the activation functions from other data points in the sequence to generate the subsequent output in a series.

A downside to DL techniques is they often require enormous data sets for training and are computationally expensive. Without appropriate tuning and validation, DL models often suffer from overfitting, especially when an adequately large data set is not available. A potential solution to circumvent this issue is through the implementation of transfer learning. In transfer learning, an algorithm is pre-trained on a larger data set before being subsequently trained on a separate but related smaller data set (i.e., the data set of interest) with the goal of achieving improved prediction performance. For instance,

a DL algorithm trained to analyze echocardiograms could first be trained on general images, where it could learn to identify common image features such as edges, shapes, and patterns and then apply that learning to characterize the echocardiographic data sets better. Another drawback of DL is that the resulting models can be more difficult to interpret than non-deep methods, as the complexity of the neural networks produces relationships that become unrecognizable. Thus, it is often advisable to begin the analysis with a more non-deep ML algorithm before moving on to DL.

Despite its nascence, DL applied to the domain of cardiology exhibits great potential (49). Although neural networks and DL have gained limited traction in risk prediction, they have been extensively utilized in image processing (13).

1.4.3.6 Topological data analysis

Topological data analysis (TDA) is a novel technique that uses the shape of data to extract meaningful insights and identify subgroups (50). This approach adopts methods of topology, a discipline of mathematics that studies robust algorithms of shape analysis, to create compact visual representations of high-dimensional data sets (51, 52). It draws on the philosophy that all data has an underlying shape and that shape has meaning. TDA amalgamates unsupervised pattern detection and network visualization by identifying and connecting data points with very similar characteristics in a multi-dimensional space and then plotting the data in a lower-dimensional space. Although it harnesses and combines several ML algorithms to elucidate the fundamental properties and the shape of complex data, TDA should be perceived as a framework for ML rather than an ML algorithm.

TDA has three key attributes that enable the extraction of pattern via shape (50): (1) coordinate freeness – TDA does not depend on the coordinate system chosen as only the applied distance function specifies the shape, (2) deformation invariance – topological properties are unchanged when the geometric shape is stretched or deformed making topology less sensitive to noise, and (3) compression – topological networks encode relationships in a simple format.

Two types of input parameters are required to construct topological models. First is the distance metric, which measures the similarity between data points. Commonly used metrics include correlation, angle, and cosine distance. Second are lenses, which are filter functions describing the distribution of the data. Multi-dimensional scaling,

variance, and principal component analysis are lenses frequently encountered in the literature. Multiple lenses can be applied in each topological model. For each lens, gain and resolution have to be defined before starting the analysis. Adjusting the resolution and gain alters the number of bins created within the range of selected lens values and the degree of overlap of these bins, respectively. Increasing the resolution will increase the number of nodes, whereas increasing the gain will increase the number of edges in the map.

After selecting a metric and one or more lenses, myriads of iterations are performed to find the most stable consensus vote for defining a "golden network" for high dimensional data shape. The generated network consists of nodes with edges between them (Figure 4). Each node represents a collection of similar data points (e.g., patients), and two nodes are connected if they have at least one data point in common. The layout of a network is chosen using a force-directed layout algorithm, and the coordinates of any individual node have no particular meaning, only the connections between them. Hence, a network can be freely rotated and translated to different positions without disrupting the interpretation of the results. Typical shapes which appear in topological networks are loops (continuous circular segments) and flares (long linear segments). Networks can be color-coded based on the outcome of interest (i.e., features not utilized during network building) to reveal clinically meaningful regions (subgroups) in the map.

In a sense, TDA is similar to cluster analysis as it also offers an efficient way of partitioning data to understand the underlying properties characterizing the subgroups within the data. However, most clustering algorithms rely on global optimization techniques, which are susceptible to noise since they consider all of the data during the optimization, whereas TDA splits data into multiple independent partitions using lens functions and runs clustering algorithms within each portion independently. TDA then combines these partial clusters into a network representation that gives an overview of the similarity between the data points. The multiple local optimizations executed by TDA dramatically reduce the effect of noise on the final results, which makes TDA suitable for constructing a connected representation of either continuous data sets or data with heterogeneous densities.

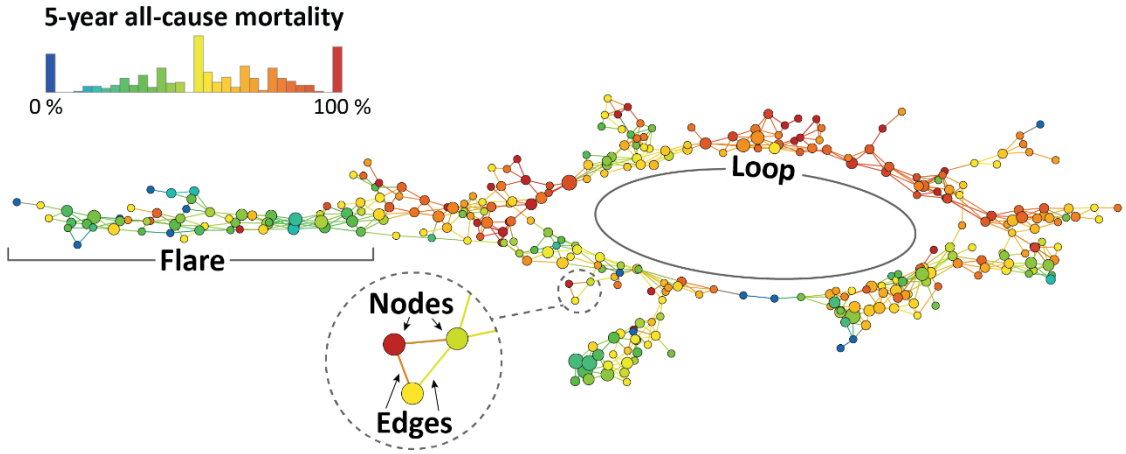


Figure 4 *Topological network of heart failure patients who underwent CRT implantation (n = 1,650)*

The topological network was created using eight pre-implantation variables (age, etiology of heart failure, type of the implanted device, QRS morphology, NYHA functional class, serum sodium, hemoglobin concentration, serum creatinine). The generated network consists of nodes with edges between them. Each node represents a collection of similar patients, and two nodes are connected if they have at least one patient in common. Typical shapes which appear in topological networks are loops (continuous circular segments) and flares (long linear segments). In this network, nodes are color-coded based on 5-year all-cause mortality. Metric: normalized correlation, lenses: 2 × multi-dimensional scaling (resolution: 20, gain: 1.40, equalized).

CRT – cardiac resynchronization therapy, NYHA – New York Heart Association

1.4.4 Model training, validation, and testing

The next step in the ML analysis pipeline is model training and evaluation. In this section, emphasis is placed on the training process of supervised ML algorithms.

Before starting the training process, data sets are typically partitioned into training and validation sets. The training set, which encompasses the bulk of all available data, is used for model derivation, and the validation set is utilized to compare different models to select the best one or to fine-tune their hyperparameters. Finally, a statistically independent test set – data that is not used in any way during training or validation, ideally originated from multiple external medical centers – can be used to assess the model's generalizability once completely optimized.

Training and validation are commonly combined through techniques such as n-fold cross-validation or bootstrapping. In these approaches, the original data set is repeatedly sampled to create training and validation sets. Then, the created sets are used recursively to fit and evaluate models. These methods are more robust than single-split validation (i.e., randomly splitting the data into a training and validation set) since they reduce the variance more efficiently. The main difference between bootstrapping and

cross-validation is that the former resamples the data with replacement (i.e., bootstrapped samples may contain multiple instances of the same object and may completely omit some of the objects), whereas the latter resamples the data without replacement. The most extensively applied form of n-fold cross-validation is 10-fold cross-validation, in which the data set is randomly divided into 10 equal folds. Ten validation experiments are then performed, with each fold used in rotation as the validation set and the remaining 9 folds as the training set. Therefore, each data point is used once for testing and nine times for training, resulting in 10 experimental ML models trained on 90% fractions. The average of results from the 10 experimental models is calculated to provide a measure of the overall performance.

1.4.4.1 The bias-variance tradeoff

During model development, we search, either stochastically or deterministically, for the best fit (i.e., the best configuration of hyperparameters). Although the searching process differs across ML techniques, each algorithm attempts to balance two competing interests: bias and variance (Figure 5). In the context of ML, bias is defined as the extent of difference between the predicted and the true values during the training phase. Variance is the sensitivity of the predictions to perturbation in the input data, and it is an indicator of generalizability. While our aim is to reduce both bias and variance, these two goals are often in conflict: decreased bias may increase variance and vica versa. For example, we can create a model that correctly predicts all adverse events in our cohort; however, this model may be configured in a way so that it captures the characteristics of the data set too specifically (i.e., it has low bias). Thus, the model would perform poorly when applied in an independent data set (i.e., it has high variance). This phenomenon is often referred to as overfitting, which is one of the most common problems encountered when using supervised ML algorithms and the main limitation to their application to real-life clinical scenarios. Several techniques exist that can reduce overfitting, such as (1) increasing the amount of training data, (2) reducing model complexity, (3) limiting the number of learning iterations, (4) balancing parameters learned in the model to obtain a simpler model that underfits on the training data but generalizes better, or (5) using an ensemble of separate models to balance out the overfitting effects of each singular model. In contrast to overfitting, underfitting is the phenomenon when the model fails to correctly

classify data or make accurate predictions during the training phase as it cannot capture the underlying patterns in the data. Techniques to reduce underfitting include (1) increasing the model complexity, (2) increasing the number of input features (e.g., by performing feature engineering), (3) removing noise from the data, and (4) increasing the duration of training to get better results.

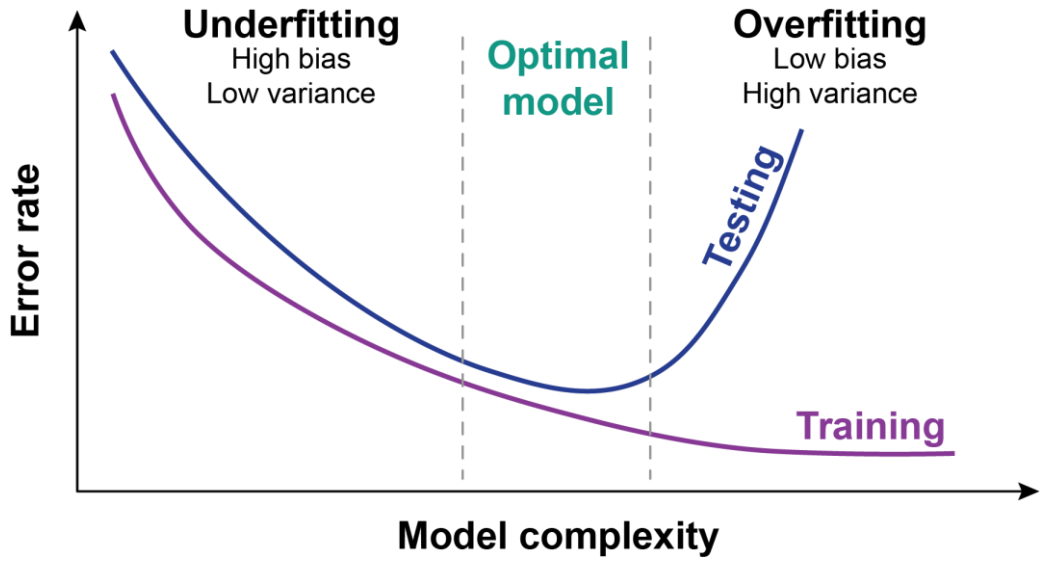


Figure 5 *The bias-variance tradeoff*

While training a machine learning model, our objective is to reduce both bias and variance. However, these two goals are often in conflict: decreased bias may increase variance and vica versa.

1.4.4.2 Performance evaluation metrics

Various evaluation metrics can be utilized to quantify the degree to which the constructed model is successfully achieving its designated task. The choice of evaluation metrics heavily impacts the ability of the model to learn from the data as it is utilized to guide all subsequent adjustments (e.g., gathering more data, fine-tuning-hyperparameters, employing regularization, or choosing an entirely different learning algorithm). Hence, depending upon the characteristics of the data set and the problem statement, the choice of an appropriate evaluation metric is essential.

1.4.4.2.1 Supervised learning – Classification

Confusion matrix is the most commonly used evaluation tool for ML-based classifiers. It is a table summarizing the occurrence of different combinations of predicted

and actual class labels, and it can be used to calculate various performance metrics. The most intuitive and convenient metric is accuracy which can be defined as the ratio of the correctly classified samples to the total number of samples in the data set. However, this metric can conceal poor model performance if the class distribution of the utilized data set is exceedingly imbalanced (e.g., predicting the occurrence of a rare disease or event). In these cases, metrics such as precision, recall, or F-score should be calculated instead. The precision for a class is the number of true positives divided by the total number of elements labeled by the model as belonging to the positive class (i.e., the sum of true and false positives). The recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e., the sum of true positives and false negatives). Simply put, high precision indicates that the model returned substantially more relevant results than irrelevant ones, whereas high recall denotes that the model returned most of the relevant results. The F-score is the harmonic mean of precision and recall and gives a better measure of the incorrectly classified cases than accuracy.

The predictive abilities of ML classifiers are often described with a receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC, often referred to as c-statistics in medical literature). As most supervised ML algorithms assign class membership probabilities to each test example, they can be used for ranking from most likely to least likely to belong to a given class. A ROC curve is created using the sensitivity and the 1 − specificity calculated at different cuts in this ranking. Optimal cut-offs can be identified using various techniques (e.g., the Youden-index), and the corresponding metrics described in the previous paragraph (i.e., accuracy, precision, recall, F-score) can be calculated. Although the AUC has become nearly universal in evaluating ML classifiers, we must be careful when there is an imbalance between the numbers of positive and negative cases, as it may inflate the true predictive capabilities. Motivated by this, another measure of performance is the area under the precision-recall curve (AUPRC). Although it is used less frequently than the ROC curve, one substantial advantage of PRC is that it can assess the model's performance more reliably when the data set is imbalanced (e.g., there is a low number of true positives).

Several other measures of performance exist for the evaluation of classifiers, such as the Matthews correlation coefficient, the logarithmic loss, or the cost curves (53).

1.4.4.2.2 Supervised learning – Regression

Mean squared error (MSE), root mean squared error (RMSE, the square root of MSE), and mean absolute error (MAE) are the three error metrics that are used most frequently for evaluating and reporting the performance of an ML-based regression model. MSE is calculated as the average squared differences between the original and the predicted values in a data set, whereas MAE is computed as the average difference between the original and the predicted values. Compared to MAE, both MSE and RMSE penalize large prediction errors more than MAE does.

Other metrics for evaluating regression models include the coefficient of determination (R^2) or adjusted R^2.

1.4.4.3 Determining the importance of predictors

Unlike traditional regression models, most ML models do not estimate an easily interpretable metric that relates the predictors to the outcome. Since they capture more complex relationships, it is generally not straightforward to summarize the relationship into any single parameter. Nevertheless, many ML algorithms attempt to summarize the impact of individual variables into metrics referred to as variable importance. The variable importance is specific to each algorithm, and its value does not generally have a causal or even statistical interpretation. Instead, the measure can often be thought of as a rank ordering of which variables are the most important to the fitted model. As each ML algorithm fits a different type of model, one would expect that different methods would come up with different (but often similar) rank orderings. Although variable importance rankings cannot replace targeted hypothesis tests of specified parameters, they can facilitate hypothesis generation and help to detect factors worthy of further study.

1.5 The potential role of machine learning in heart failure

HF is a clinical syndrome characterized by dyspnea, fatigue, and clinical signs of pulmonary or systemic congestion leading to frequent hospitalizations, impaired quality of life, and decreased life expectancy (54). It represents a global healthcare issue affecting 1-2% of the adult population in developed countries, rising to above 10% among patients >70 years of age (55-57). Despite the continuous advancement and increasing availability

of therapeutic agents and medical devices, outcomes still remain unsatisfactory in this patient population.

Due to its favorable attributes described above, ML has a true potential to revolutionize HF care and ultimately improve outcomes in these patients (Figure 6). ML has been leveraged in an attempt to facilitate the diagnosis of HF, the classification of HF patients into subgroups requiring different treatment strategies, and the prognostication of HF patients in order to deliver more tailored care (17, 58). ML techniques can be applied in other aspects of the management of HF as well. For instance, they can predict whether the patient will adhere to the prescribed medications (59, 60) or identify HF patients at risk of depression (61). ML can also be employed to predict LV filling pressures (62) or to protect implantable devices from cyberattacks (63). Although these are only cherry-picked use cases, they make it apparent that ML can be successfully applied for countless prediction tasks related to HF.

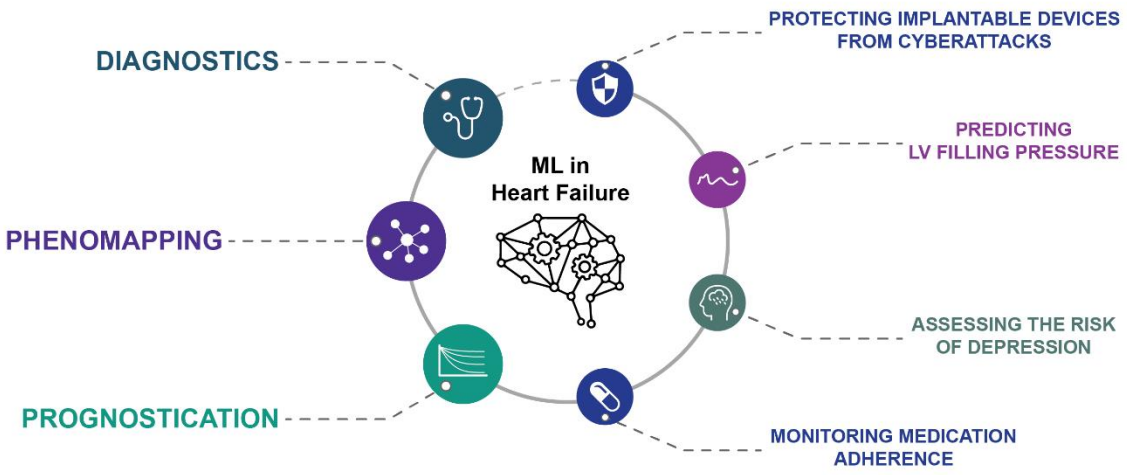


Figure 6 *Potential role of ML in heart failure*
LV – left ventricular, ML – machine learning

1.5.1 Diagnosis of heart failure

Currently, the diagnosis of HF relies upon patients' history, physical examination, and both laboratory and imaging data (56). ML-based methods aim to improve diagnosis through leveraging data found from each of these sources, including but not limited to ECG, echocardiography, and EHR data (17, 58). The accurate and early diagnosis of HF

via ML techniques will allow treatments and interventions to be delivered in a more efficient and targeted way and may decrease medical costs as well.

The importance of ECG in the workup of HF is well established; however, it is primarily utilized to identify the underlying etiologies or complicating factors, such as arrhythmias (56). As HF-related pathophysiological alterations might be reflected in the electrical activity of the heart, many research groups have hypothesized that the ML-based analysis of ECG signals will enable the accurate detection of HF. Such an approach would facilitate screening for HF in an outpatient setting and in the emergency department as well. Recently, Attia et al. trained and evaluated a CNN using ECG data to identify patients with asymptomatic left ventricular (LV) dysfunction (64). Their ML model predicted the presence of systolic dysfunction (defined as ejection fraction [EF] $\leq$35% by echocardiography) with an AUC of 0.93. When the performance of this algorithm was validated prospectively, it was able to detect LVEF of $\leq$35% with an AUC of 0.92 (65). Moreover, the algorithm was tested in an external cohort of patients presenting with dyspnea to the emergency department and showed an AUC of 0.89 (66). Kwon et al. performed a similar but slightly different analysis to detect HF with reduced EF (HFrEF) based on ECG features (67). For the detection of HFrEF, their deep neural network exhibited AUCs of 0.84 and 0.89 in the internal and external validation sets, respectively, and it outperformed both logistic regression and random forest. Importantly, not only HFrEF but also HF with preserved EF (HFpEF) or LV diastolic dysfunction can be reliably predicted from ECG signals with the help of ML algorithms (68, 69).

Echocardiography has been the cornerstone of HF characterization and management for decades (56). Nevertheless, the interpretation of echocardiograms requires a skilled reader, and measurements are subject to inter-reader variability. To mitigate the effect of this factor, ML algorithms can be utilized to perform an automatic and objective evaluation of echocardiograms. To this end, Zhang et al. implemented a CNN which was capable of identifying cardiac views with an accuracy of 84% and calculating LVEF and longitudinal strain with median absolute differences of 6.0% and 1.4% when compared with manual tracings, respectively (70). Ouyang et al. reported another interesting approach, in which they trained a CNN on full-length echocardiographic videos containing more cardiac cycles instead of manually curated still images at end-systole and end-diastole (71). Their model was capable of predicting EF

with a mean absolute error of 4.1% and 6.0% (when compared with manual tracings) and reliably identifying HFrEF with AUCs of 0.97 and 0.96 in an internal and external test set, respectively. Although these two methods have notable differences, both rely on image segmentation (i.e., detecting the endocardial border and the LV blood pool). In contrast, Asch et al. proposed a DL-based model which estimates EF using still echocardiographic images without performing image segmentation and volume measurements (72). Their model achieved an accuracy of 92% for the detection LVEF ≤35%. Of note, instead of raw image data, standard echocardiographic variables or velocity, strain, and strain rate traces can also be used to train ML algorithms to diagnose HF, including HFpEF (73-75).

EHR systems are an integral part of everyday clinical practice storing an immense and ever-growing amount of data, mainly in an unstructured fashion. By supplying that data to ML models, we can substantially improve the identification of patients with HF, particularly of those without a prior clinical diagnosis. Blecker et al. employed an L1-regularization logistic regression model using unstructured (i.e., free-text notes and procedure reports) and structured data elements to identify patients with HF while being hospitalized (76). With an AUC of 0.97, this model exhibited better performance than the one utilizing structured data only, suggesting that also unstructured data contain non-negligible information. Similar results were reported in another study of the same research group, in which L1-regularization logistic regression models utilizing unstructured medical notes achieved AUCs of 0.99 to identify hospitalizations with a principal discharge diagnosis of acute decompensated HF (77). Besides enhancing the diagnosis of HF, ML-based analysis of EHRs can also be used to automatically extract other pieces of clinically relevant information from medical notes, such as New York Heart Association (NYHA) functional class (78), indicators of LV function, prescribed HF medications, or reasons for not prescribing a given medication (79).

Another potential use case of ML algorithms is to forecast the future onset of HF. Ambale-Venktesh et al. used a random survival forest algorithm to predict cardiovascular outcomes, including incident HF, among patients enrolled in the MESA (Multi-Ethnic Study of Atherosclerosis) (80). They identified the most relevant predictors for each outcome included in the study, which for the prediction of HF onset resulted in an AUC of 0.84, a modest improvement over the Cox proportional hazards model-based MESA-

HF risk score (AUC of 0.80) (81). Segar et al. sought to predict incident HF hospitalization in patients with type 2 diabetes mellitus (82). They trained and validated a random survival forest model, which appeared superior to the best performing Cox-based method. Choi et al. utilized a more sophisticated algorithm – RNN models – to leverage temporal relations among time-stamped events for the detection of incident HF (83). With observation windows of 12 and 18 months, the AUCs of the RNN models were respectively 0.78 and 0.88 for predicting the initial diagnosis of HF, which were notably higher compared to the performance of conventional ML methods such as regularized logistic regression, SVM, and KNN classifiers. Using longitudinal EHR data, Chen et al. assessed the performance of three ML models (L1-regularized logistic regression, random forest, and RNN) trained to detect a future diagnosis of HF in primary care patients (84). RNN model performance was superior to the other two methods under various conditions, including when data were less diverse, the prediction window length decreased, the data density increased, or the observation window length increased. The potential utility of RNN-based analysis of EHR data sets was also demonstrated by Rasmy et al., whose RNN outperformed L2-regularized logistic regression in predicting the risk of HF onset (AUC: 0.82 vs. 0.79) (85).

Beyond the applications mentioned above, ML has been coupled with other less usual data sources to improve the diagnosis of HF. Nirschl et al. trained a CNN model using hematoxylin and eosin-stained images of cardiac tissue to diagnose HF in postmortem patients (86). Interestingly, their model had an AUC of 0.99 and outperformed two experienced clinical pathologists as well. Liu et al. analyzed digital heart sound recordings to discriminate between HFpEF patients and healthy volunteers with no history of cardiovascular disease (87). Although they used a relatively simple feed-forward network with a single hidden layer, their model still exhibited an accuracy of 96% to identify HFpEF patients.

1.5.2 Identification of heart failure subtypes

Traditionally, HF patients have been stratified according to LVEF (56). A wide variety of effective treatments has been found for patients classified as having HFrEF (56), although the same success has evaded researchers investigating the therapeutic options for HFpEF (88-91). This is partly attributable to the fact that HFpEF is a

heterogeneous clinical syndrome (92, 93). Thus, defining therapeutically homogeneous subclasses of HFpEF patients would be essential, which represents an excellent opportunity for unsupervised ML algorithms (i.e., clustering) to shine.

Shah et al. were one of the firsts to apply unsupervised algorithms (namely hierarchical and model-based clustering) to cluster patients into three unique HFpEF phenogroups that differed markedly in clinical characteristics, cardiac structure/function, invasive hemodynamics, and outcomes (94). They also tested the validity of the three phenogroups in a separate patient cohort and found similar outcomes after matching patients to the newly defined groups. More recently, using data from TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial), Cohen et al. performed latent-class analysis to identify HFpEF phenogroups based on standard clinical features and assessed differences in multiple plasma biomarkers, cardiac and arterial structure/function, prognosis, and response to spironolactone (95). The subgrouping of HFpEF has also been attempted by Hedman et al., aiming to gain insights into the underlying proteomics profiles of each cluster (96). By performing model-based clustering using echocardiographic, clinical, and standard laboratory variables collected from stable HFpEF outpatients, they identified six groups with significant differences in outcomes and levels of inflammatory and cardiovascular proteins. Not only resting but also post-exercise imaging data can be incorporated into cluster analysis for the more profound characterization of HFpEF subgroups, as demonstrated by Przewlocka-Kosmala and colleagues (97). In their study, they applied hierarchical clustering to define two distinct phenotypes of patients with symptomatic HFpEF: (1) those with reduced and (2) those with preserved chronotropic and/or diastolic reserve, with the latter group showing a lower incidence of cardiovascular hospitalization or death during the 2-year follow-up period.

Similar studies have been performed in patients with HFrEF and acute decompensated HF (98, 99). Each of these studies delineated unique subgroups of patients with graduated outcomes depending on cluster assignment; nevertheless, the clinical utility of these clusters in each instance has yet to be established, and it should also be investigated whether changes in outcomes can be achieved by tailoring therapies or prevention strategies to individual groups.

Although the studies discussed above focused on the phenomapping of either HFpEF or HFrEF, clustering can also be conducted using data from HF patients across the entire EF spectrum. Motivated by this idea, Ahmad et al. performed k-means clustering using data from the Swedish Heart Failure Registry in order to identify subgroups of HF patients with distinct phenotypes and survival independent of LVEF (100). Interestingly, in terms of prognostication, assignment to the newly identified clusters proved to be superior to the current EF-based stratification of HF patients (AUC of 0.68 vs. 0.52). In addition, they reported that response to pharmacotherapies (diuretics, angiotensin-converting enzyme inhibitors, β-blockers, and nitrates) differed across clusters. These results suggest that it may be possible to apply ML algorithms to pinpoint individuals who will benefit from a specific therapy.

Of note, phenomapping does not have to be restricted to HF patients as it can be implemented to explore the entire spectrum of cardiac function, from normal to end-stage HF. An excellent example of this concept was provided by Cho et al., who used TDA to integrate multiple echocardiographic variables and visualize the distinct phenotypes of cardiac function (101). In the generated topological network, four phenogroups (i.e., clusters) were identified, which differed significantly in clinical characteristics, cardiac structure and function, hemodynamics, and outcomes, and showed meaningful correlations with the clinical stages of HF. Finally, a deep neural network was trained using the newly identified phenogroups as class labels, and it was able to classify patients into one of the four clusters with an AUC of >0.80.

1.5.3 Prognostication of heart failure patients

In addition to improving the diagnosis and phenogrouping of HF, ML techniques can be leveraged to accurately predict different outcomes.

1.5.3.1 Predicting 30-day hospital readmission

Readmission within 30 days after an HF hospitalization is recognized both as an indicator for disease progression and a source of a considerable financial burden to the healthcare system. Consequently, the identification of patients at risk for 30-day readmission is a key step in meeting the individual needs of a patient while optimizing the distribution of healthcare resources among all patients.

Using a vast amount of EHR data, Golas et al. trained deep unified networks – a new mesh-like network structure of DL designed to avoid over-fitting – to predict 30-day all-cause readmission of HF patients (102). In the 10-fold cross-validation procedure, deep unified networks were found to be more powerful than both gradient boosting (mean AUCs of 0.71 vs. 0.65, respectively) and logistic regression (mean AUCs of 0.71 vs. 0.66, respectively). The usefulness of neural networks in this prediction task was demonstrated by Awan et al. as well, who implemented a multi-layer perceptron (i.e., a feed-forward ANN) to predict 30-day HF readmission (103). Their model showed an improved performance to predict 30-day HF readmission compared to logistic regression (AUCs of 0.62 vs. 0.58, respectively). Another example for the successful application of ANN was provided by Liu et al., who achieved an AUC of 0.64 to predict 30-day readmission of HF patients using an ANN with medical code embedding (104).

Besides neural networks, other ML algorithms have also exhibited encouraging results in the prediction of 30-day readmission. Mortazavi et al. explored the performance of conventional ML algorithms for the prediction of both all-cause and HF-only readmissions after hospitalization for HF (105). In the 30-day all-cause readmission prediction, random forest exhibited the best performance and provided an 18% improvement in discrimination over logistic regression (AUC: 0.63 and 0.53, respectively), whereas, for readmissions due to HF, gradient boosting achieved the most substantial improvement over logistic regression (AUC: 0.68 and 0.54, respectively). In another study, Yu et al. utilized an SVM leveraging patient data acquired at both admission and discharge during the index hospitalization to predict 30-day readmission due to HF (106). According to their findings, the SVM model was capable of outperforming the LACE index for readmission (AUC: 0.65 vs. AUC: 0.56).

1.5.3.2 Predicting mortality

In the last decades, predicting the mortality of HF patients has attracted increased attention, and significant strides have been achieved with the application of ML.

In a study conducted by Samad et al., a random forest model including a large panel of echocardiographic features yielded the highest prediction power among the evaluated ML classifiers and outperformed the Seattle Heart Failure Model (SHFM) in the 5-year survival prediction (AUC: 0.80 vs. 0.63) (107). In the same year, Adler et al.

created an ML-based risk score – the MARKER-HF risk model – to discriminate HF patients with high and low mortality risk (108). Their model was found to have a superior ability to prognosticate HF patients compared to pre-existing, non-ML-based risk scores (AUC: 0.88 vs. 0.78 [the Intermountain Risk Score], 0.74 [the Get With The Guidelines-Heart Failure risk score – GWGT-HF], 0.63 [the Acute Decompensated Heart Failure Registry score]). For the prediction of 1-year outcomes, Hearn et al. implemented a feed-forward neural network incorporating breath-by-breath data from cardiopulmonary exercise testing (109). Their model achieved an AUC of 0.84 to predict clinical deterioration (i.e., initiation of mechanical circulatory support, listing for heart transplantation or mortality) within 1-year following cardiopulmonary exercise testing. Kwon et al. proposed an echocardiography-based deep learning model which predicted in-hospital, 1- and 3-year mortality among acute HF patients with AUCs of 0.88, 0.78, and 0.81, respectively (110). In addition, the proposed model outperformed the GWGT-HF score in the prediction of in-hospital mortality (AUC: 0.88 vs. 0.73), the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) score in the prediction of 1- and 3-year mortality (AUC: 0.78 vs. 0.72 and 0.81 vs. 0.73, respectively). Another group of researchers developed and validated contrast pattern-aided logistic regression-based models to predict 1-, 2-, and 5-year survival in HF using data from EHRs (111). In the validation cohort, the model exhibited AUCs of 0.94, 0.83, and 0.79 to predict 1-, 2- and 5-year, respectively, outperforming random forest, adaptive boosting, SVM, and logistic regression. The inspiring results of these studies support the utilization of ML-based approaches for the prognostication of patients with HF and in other settings where risk prediction has been challenging so far.

1.5.3.3 Predicting response to left ventricular assist device

LV assist device (LVAD) has rapidly emerged as a durable and safe therapy for end-stage HF patients (112). Though originally conceived for bridge-to-transplant indication, advancements in medical management and technology with the arrival of newer generation devices have significantly improved patient outcomes, leading to its increasing use as destination therapy (113, 114). However, the response to LVAD is still varying, and conventional statistics-based risk prediction tools for adverse outcomes such as right ventricular (RV) failure and survival have limited predictive power. Thus, ML-

based solutions to enhance the prognostication of these patients have been investigated intensively in the last decade.

Using multiple pre-LVAD variables of a large patient cohort from the INTERMACS (Inter-Agency Registry for Mechanically Assisted Circulatory Support), Kanwar et al. implemented NB classifiers to predict 1-, 3-, and 12-month survival following LVAD implantation (115). Their models achieved AUCs of 0.70 and 0.71 in 90-day and 1-year survival prediction, respectively, and outperformed the HeartMate II Risk Score (AUC: 0.61 and 0.59, respectively) (116). Moreover, in another study conducted by the same group of researchers, the performance of similar ML models was evaluated in the prediction of acute (<48h), early (48h-14days), and late (>14days) RV failure (117). The AUCs of the constructed NB models were above 0.83, and they outperformed the previously published risk scores for RV failure, including the RV failure risk score (118) and the Drakos score (119). All these findings imply that ML models have the potential to play a considerable role in clinical decision-making while screening candidates for LVAD therapy.

1.5.3.4 Predicting response to CRT

Cardiac resynchronization therapy (CRT) is a key component in the management of symptomatic HF with reduced LVEF and wide QRS complex (57). Although CRT improves mortality, functional capacity, clinical symptoms, and quality of life in a certain patient subpopulation, not everyone benefits equally, and mortality rates remain high among these patients (120-124). The recognition of this variability in outcomes has prompted efforts in the response prediction and risk stratification of CRT patients based on pre-implantation assessments. However, the currently available, conventional statistics-based risk scores have several shortcomings (e.g., lack of generalizability and impact analysis, omitting routinely assessed, powerful predictors), which hamper their utilization in everyday clinical practice (125). Therefore, more precise and personalized methods are required, and ML seems to be a promising tool to meet this compelling demand (11).

To this end, Cikes et al. studied HF patients from the MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy) to identify responders to CRT-defibrillator (CRT-D) using multiple kernel learning and k-

means clustering (126). Four phenogroups were delineated that significantly differed in pre-implantation clinical characteristics, biomarker values, measures of LV and RV structure and function, and in the extent of volumetric response. In addition, two phenogroups were found to be associated with a substantially better treatment effect of CRT-D on the primary outcome (all-cause death or non-fatal HF event) than the other two clusters. Another group of researchers employed a random forest-based model to predict all-cause mortality or HF hospitalization at 12 months following CRT implantation using data acquired in the COMPANION (Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure) trial (127). Their model achieved an AUC of 0.74, and it also discriminated the risk of the composite endpoint of all-cause death or HF hospitalization better than subgroups created based on left bundle branch block (LBBB) morphology and QRS duration.

More recently, the capabilities of ML have also been demonstrated using the data of the SMART-AV (SMARTDELAY-determined AV optimization) trial by Howell et al., who applied ML to predict short-term CRT response (128). In their series of experiments, the best performing model – an adaptive LASSO model – was able to predict the composite endpoint (freedom from death / HF hospitalization and a >15% reduction in LV end-systolic volume index at 6-month follow-up) with an AUC of 0.76 suggesting that ML-based tools may facilitate early post-implantation care planning that will ultimately improve long-term CRT outcomes.

ML has been successfully applied not only on data from multi-centric, prospective trials but also in retrospective registries to predict CRT response. Feeny et al. analyzed CRT patients using an NB classifier to identify responders to CRT (defined as ≥10% increase in LVEF) and predict survival (129). In the test cohort, the implemented NB classifier improved both response classification (AUC: 0.70 vs. 0.65) and survival prediction (Harrell's concordance index: 0.61 vs. 0.56) compared to the 2013 American College of Cardiology (ACC) / American Heart Association (AHA) HF guidelines (130). In another study, the same research group used principal component analysis and k-means clustering to analyze the QRS patterns of patients prior to CRT implantation (131). They identified two clusters with prominent differences in the risk of reaching the composite endpoint (death, LVAD implantation, or heart transplantation) and the extent of mean LVEF improvement. Moreover, using the combination of cluster membership and QRS

area, they could also pinpoint patients with LBBB and QRS duration <150 milliseconds who had similar outcomes to those with LBBB and QRS duration ≥150 milliseconds, and identified patients with LBBB having outcomes similar to non-LBBB patients. Thus, the ML-based stratification of LBBB may represent an objective method for candidate selection without requiring subjective but strict definition for LBBB.

Cai et al. have also demonstrated that ML might improve upon clinical guidelines (132). In their recently published study, the proposed ML model using the combination of clinical and ECG features predicted >5% improvement in LVEF at 6-month follow-up with an AUC of 0.76, outperforming models using only QRS > 120 ms, LBBB, right bundle branch block, and intraventricular conduction delay (AUC: 0.56). Accordingly, their risk stratification tool identified responders more efficiently than clinical guidelines, suggesting that ML can enhance the care management of HF patients by facilitating the identification of high-risk patients.

ML can also be employed to explore the most important predictors of volumetric response and other clinical endpoints in patients undergoing CRT implantation. An excellent example was provided by Galli et al., who used a feature selection method called the Boruta algorithm to identify the key clinical, echocardiographic, and electrocardiographic determinants of volumetric response and outcomes (133). Septal flash, apical rocking, and ischemic heart disease were the most important variables associated with volumetric response (defined as ≥15% decrease in LV end-systolic volume), underscoring the relevance of visual assessment of LV mechanical discoordination and ischemic cardiomyopathy. Tricuspid annular plane systolic excursion was found to be the most important feature associated with outcomes (i.e., the composite of heart transplantation, LVAD implantation, or all-cause death during follow-up), followed by LV size, NYHA functional class, renal function, fractional area change, LVEF, and RV free wall longitudinal strain. Moreover, the random forest model using exclusively the selected features achieved outstanding performance in the prediction of both volumetric response (AUC: 0.81) and outcomes (AUC: 0.84).

Clinical notes represent another important yet often untapped data source that ML algorithms can leverage to predict outcomes. Inspired by this idea, Hu et al. trained gradient boosting classifiers that utilized structured data (e.g., laboratory results, medications, billing codes) and two-word features (i.e., bigrams) extracted from clinical

notes via natural language processing (134). Their final model was capable of identifying patients who were unlikely to benefit from CRT (<0% improvement in LVEF or death by 18 months) with an AUC of 0.75.

As demonstrated by Jing et al., ML can also be utilized to priority-rank HF patients based on the predicted degree of benefit from closing care gaps, such as implantation of a CRT device or optimization of pharmacotherapy (135). Using a gradient boosting model capable of predicting 1-year all-cause mortality with an AUC of 0.77, a subgroup of HF patients with high risk of mortality could be delineated who had high benefit (>10% reduction in mortality rate) from closing care gaps. However, not all high-risk patients were predicted to have high benefit, as evidenced by a subgroup of patients with high baseline risk but minimal risk reduction after closing the care gaps. Moreover, prioritizing patients for intervention based on the predicted reduction in 1-year mortality risk outperformed other priority rankings, such as random selection or the Seattle Heart Failure risk score. Thus, in a resource-constrained environment, ML has a potential value in optimizing the distribution of healthcare resources and maximizing the total number of lives saved.

Seeing the promising results of these studies, we may deduce that ML has the ability to effectively risk-stratify patients undergoing CRT implantation and identify responders; thus, it may facilitate candidate selection and optimization of post-implantation care. Beyond these potential areas of application, ML might be applied for various additional CRT-related tasks, such as for protecting cardiovascular implantable electronic devices (CIEDs) from cyberattacks (63), for the rapid and automated identification of the manufacturer and the model group of CIEDs based on chest radiographs (136-138), or for analyzing other data captured by CRT devices such as heart rate variability and pulmonary impedance (139, 140).

2. OBJECTIVES

1. Applying topological data analysis to integrate echocardiographic features of left ventricular structure and function into a patient similarity network

In cardiovascular medicine, one of the major research priorities is to prevent adverse clinical events and hospitalization by risk factor management and by earlier detection of subclinical cardiac dysfunction. This research has led to a plethora of imaging approaches with diverse technical underpinnings to assess various and often overlapping aspects of cardiac function. A diagnostic imaging protocol can produce numerous parameters, each with its strengths and limitations (141). However, due to the lack of unanimity on the combination and use of these parameters to depict a single patient or a group of patients with similar characteristics, there is a paramount need to develop a staging method that can integrate multiple tests and diagnostic variables at the point of care. Accordingly, we applied TDA-based network analysis to detect patient similarity patterns using a cross-sectional multi-parametric echocardiographic data set. We subsequently investigated the prognostic value of the topological network and explored whether the longitudinal course of the disease could be tracked along the topological map to assess the risk of cardiac events in an index patient.

2. Designing and evaluating a machine learning-based risk stratification system to predict all-cause mortality of patients undergoing CRT implantation

CRT has been shown to decrease morbidity and mortality in appropriately selected patients, yet some patients show limited or no response to therapy (142, 143). Since this discovery, interest has spiked in attempting to accurately risk-stratify this patient population; however, conventional statistics-based risk scores have achieved only modest performance. Therefore, we sought to design and evaluate an ML-based risk stratification system to predict 1-, 2-, 3-, 4-, and 5-year mortality from pre-implantation parameters of patients undergoing CRT implantation. We hypothesized that ML can capture high-dimensional, non-linear relationships among clinical features, and a risk stratification system can be developed that predicts mortality for individual patients more accurately than the currently available risk scores.

3. Exploring the sex-specific differences and similarities in the predictors of mortality among patients undergoing CRT implantation

In HFrEF, several studies have highlighted sex-related differences that involve multiple aspects of the syndrome, such as epidemiology, pathophysiology, phenotyping, and prognosis (144). While women with HFrEF have better survival and lower hospitalization rates, they have a greater burden of symptoms and more impaired health-related quality of life than men (145). Although sex disparities are also remarkable in the accessibility to HF device therapy, including CRT (146-148), women are more likely to respond favorably and derive a greater survival benefit from CRT implantation (149-152). Nonetheless, the sex-related differences in both short- and long-term outcomes and the varying importance of different predictors are still scarcely explored in this patient population (153). Motivated by this gap in knowledge, we aimed to explore the sex-specific differences and similarities in the predictors of mortality using advanced ML-based approaches.

3. METHODS

3.1 Study population and methods of investigating interpatient similarities using topological data analysis

3.1.1 Outline of the study protocol

Our study consisted of two parts. First, we used a primary cohort (including patients from a retrospective study and two prospective registries) to create a cross-sectional representation of patients across different stages of cardiac disease. After we developed a topological network from the retrospective study, we added the data from the prospective registries to validate the persistence and stability of network topology. Second, we tested whether we could perform personalized predictions for a previously unseen group of patients (secondary cohort) who had two echocardiographic evaluations. As these patients underwent two echocardiographic examinations, the change in their location on the network was also monitored to investigate whether the network could represent the changes in cardiac disease staging.

The institutional review board approved the study protocol (approval No. 1706617714 and 1706632601), and all study participants in the prospective studies provided written informed consent.

3.1.2 Study population

A flowchart of the study population, inclusion and exclusion criteria are provided in Figure 7. The retrospective group included a convenience sample of 866 outpatients (65 ± 17 years, 44.7% males) in sinus rhythm who were referred for echocardiographic assessment of cardiac function between March 2013 and December 2015 at the Icahn School of Medicine at Mount Sinai (New York, New York). The prospective groups included 468 patients (55 ± 15 years, 41.7% males) enrolled between July 2017 to February 2018 in two ongoing patient registries at West Virginia University (Morgantown, West Virginia) that followed two prospective trials (Analysis of Surface ECG Signals to Identify Myocardial Dysfunction in Patients at Risk for Coronary Artery Disease, NCT02560168; and Evaluation of Cardiopulmonary Diseases by Ultrasound, NCT02248831). The pooled patients from the retrospective study and the two prospective

registries formed the primary cohort, which was used for developing the patient-patient similarity network.

For personalized patient predictions, we tested the topological model in 96 additional patients (secondary cohort, 58 ± 15 years, 51.0% males) who had two consecutive echocardiographic examinations. Follow-up data for this cohort were collected after the second echocardiographic assessment.

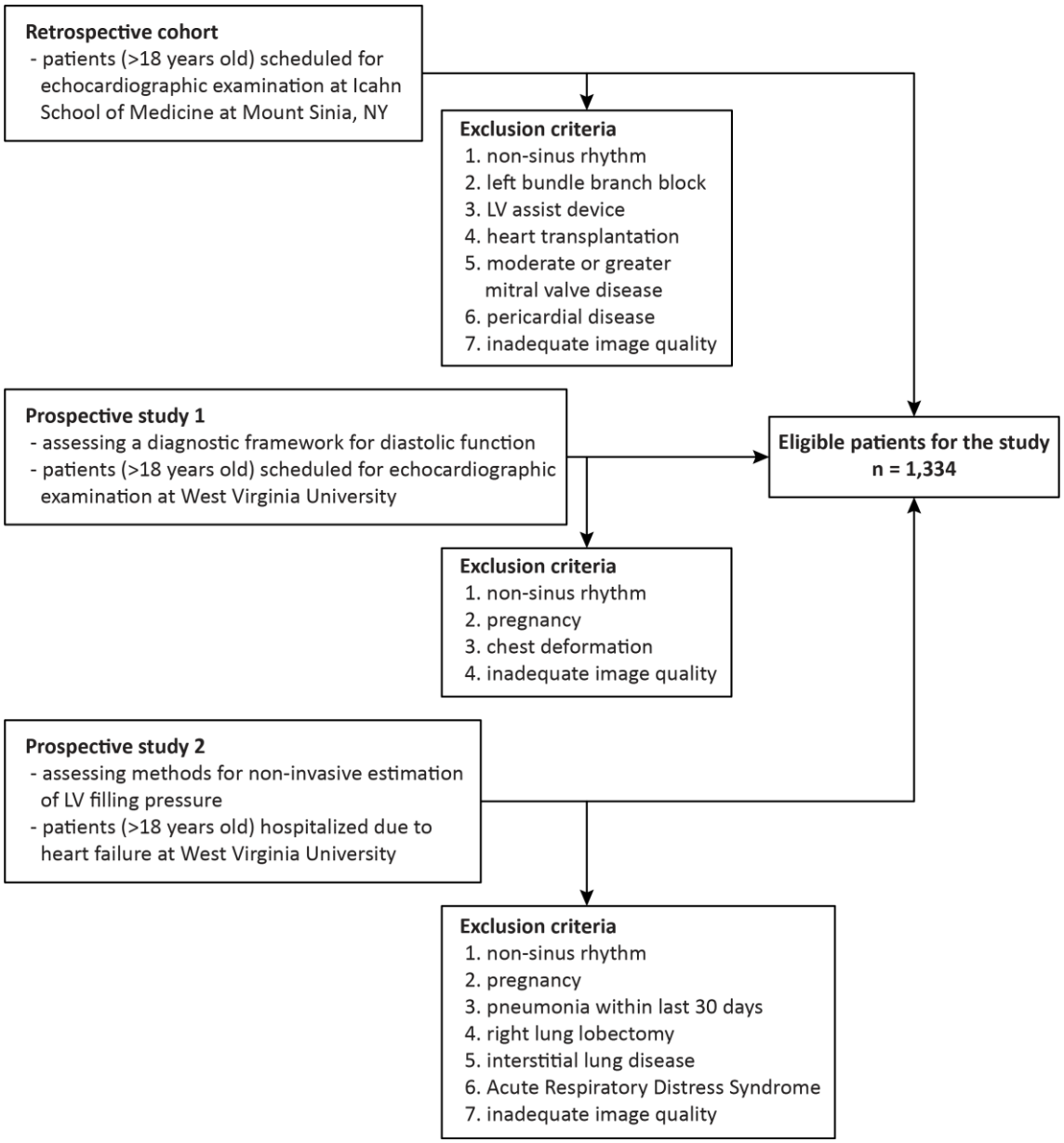


Figure 7 *Flowchart of the study protocol (154)*

LV – left ventricular

3.1.3 Clinical characteristics

Hypertension was defined as systolic blood pressure >140 mmHg or diastolic blood pressure >90 mmHg, a physician-documented history of high blood pressure, or use of any antihypertensive medication. Hyperlipidemia was defined by the presence of a physician-documented history of hyperlipidemia, the use of any lipid-lowering agents, or an abnormal lipid panel (total cholesterol >200 mg/dL, low-density lipoprotein cholesterol >100 mg/dL, triglycerides >150 mg/dL, high-density lipoprotein <40 mg/dL). Diabetes mellitus was defined as the presence of a physician-documented history of diabetes, use of any oral hypoglycemic agents or insulin, glycated hemoglobin >6.5%, or non-fasting blood glucose level >200 mg/dL. Among the prospectively enrolled patients, NYHA functional classification and ACC/AHA HF stages were also assessed.

3.1.4 Echocardiography

A standard acquisition protocol consisting of loops from parasternal, apical, and subxiphoid views was used according to current guidelines (141, 155). After measuring the LV posterior wall diameter, ventricular septal diameter, and LV internal diameter in end-diastole, LV mass was calculated and indexed to body surface area. On apical two- and four-chamber loops, the LV end-diastolic volume, end-systolic volume, and LVEF were calculated using the biplane Simpson's method of discs, and the left atrial volume was computed using the biplane area-length method. Early (E) and late (A) diastolic transmitral flow velocities were obtained from the apical four-chamber view using pulsed-wave Doppler. Spectral pulsed-wave tissue Doppler-derived early diastolic relaxation velocity (e') at the septal mitral annular position was acquired in the apical four-chamber view. The ratios of E/A and E/e' were calculated as Doppler echocardiographic estimates of the LV filling pressure. Continuous-wave Doppler was applied on the tricuspid valve in different windows (apical four-chamber, parasternal RV inflow, parasternal short-axis views) to record the tricuspid regurgitation signal, and tricuspid regurgitation peak velocity (TRV) was measured as the highest value recorded from all views. All measurements were obtained in at least three consecutive cardiac cycles, and average values were used in the final analysis.

3.1.5 Topological data analysis

We applied TDA to detect patterns in multi-dimensional echocardiographic parameters by studying the geometrical structure obtained from the network that signified a compressed representation of high-dimensional data for patient similarity analysis (50). The notion of expressing the shape of data using TDA was extensively validated and successfully applied to different areas of health sciences (156-162). In the created topological network, each node represented a cluster of patients, whereas edges connected nodes that had at least one patient in common. Nodes were color-coded based on the average value of the parameter of interest (e.g., LVEF or LV mass). The nodes were colored red for the most extreme abnormal values, whereas they were colored blue for the maximum normal value. There was a gamut of colors for the average values in between.

TDA was performed using the cloud-based Ayasdi Workbench (version 7.4, Ayasdi Inc., Menlo Park, CA). Nine echocardiographic parameters were used to create the topological network, namely, LVEF, LV mass index, E, A, E/A ratio, e', E/ e' ratio, left atrial volume index, and TRV. The steps in generating the TDA network are provided in Figure 8. Data were analyzed using a normalized correlation metric (Equations 3 and 4) with two multi-dimensional scaling lenses (resolution: 66, gain: 2.7, equalized).

$$NormCorr(X,Y) = 1 - r(X',Y')$$

where X', Y' are the mean-centered and variance-normalized X and Y:

$$r(X,Y) = \frac{N\sum_{i=1}^{N} X_i Y_i - \sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{\sqrt{N\sum_i X_i^2 - (\sum_i X_i)^2}\sqrt{N\sum_i Y_i^2 - (\sum_i Y_i)^2}}$$

Equations 3 and 4 *Normalized correlation*

Once the TDA-based loop structure was created, we divided the loop into four regions for further analyses. To this end, we calculated the multi-dimensional Euclidean distance (Equation 5) of the input features from the guideline-defined cut-off values for each patient. The distances were then used in the network for generating auto-groups using agglomerative hierarchical clustering, which identified 40 clusters of nodes within the loop. The mean distance for each cluster was calculated, and the smallest distance was identified. Subsequently, 9 clusters adjacent to this cluster were selected to create the first

region in the loop, and the remaining 30 clusters were divided into three regions containing 10 adjacent clusters of nodes each.

$$d_{x,y} = \sqrt{\sum_{j=1}^{J} (x_j - y_j)^2}$$

Equation 5 *Euclidean distance formula generalized to higher-order dimensions*
x and y are two J dimensional vectors

After the network was created in the primary cohort (containing retrospective and prospective cohorts), we trained a random forest-based classifier using the echocardiographic data of the primary cohort to predict the region that each patient from the secondary cohort might belong to. This allowed us to predict the characteristics of the patients and the outcomes that a given patient might have experienced.

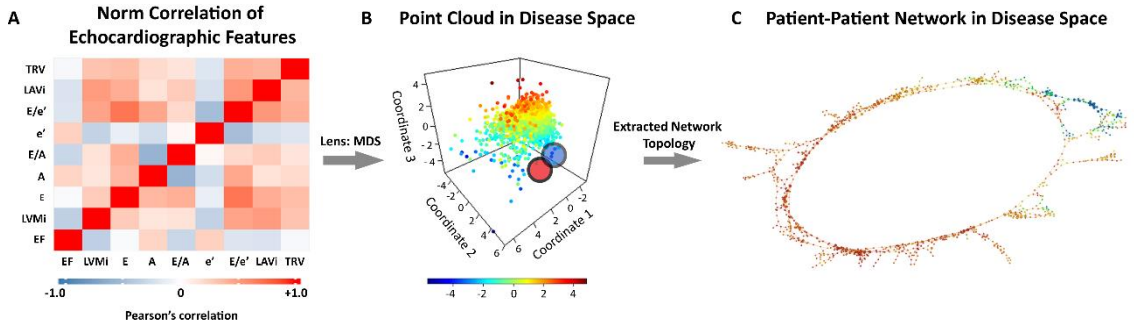


Figure 8 *Steps of topological data analysis (154)*

(A) Our data set containing echocardiographic features was analyzed using TDA from a bivariate correlation matrix. (B) Data were processed with two multi-dimensional scaling lenses to generate the disease space. TDA resamples the disease space multiple times to identify similar patients and links them to nodes (red and blue circles). (C) The patient-patient network was created to provide a simple visual representation of the data.

A – late diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, E – early diastolic transmitral flow velocity, E/A – early to late diastolic transmitral flow velocity ratio, E/e' – early diastolic transmitral flow to annular velocity ratio, EF – ejection fraction, LAVi – left atrial volume index, LVMi – left ventricular mass index, TDA – topological data analysis, TRV – tricuspid regurgitation peak volume

3.1.6 Clinical outcomes, endpoints, and staging

Patient electronic medical records were reviewed for post-echocardiographic follow-up. Hospitalizations were classified based on the International Classification of Diseases-10th Revision coding. Endpoints were death from a major adverse cardiac event (MACE, defined as myocardial infarction, acute coronary syndrome, acute decompensated HF, cardiac arrest, or arrhythmia) and first MACE-related rehospitalization. The time to each endpoint was measured from the date of the echocardiographic examination used in the study. Clinical cardiac disease staging was performed using NYHA functional class assessment, ACC/AHA HF staging, and the MAGGIC risk score. The ability of the MAGGIC score to predict death and cardiovascular hospitalization events related to MACE was well-validated (163, 164).

3.1.7 Statistical analysis

Continuous variables are expressed as median (interquartile range), while categorical variables are reported as frequencies and percentages. Between-group comparisons were performed using Chi-squared test or Fisher's exact tes (for categorical variables), and analysis of variance (ANOVA), Kruskal-Wallis test, or Kolmogorov-Smirnov test (for continuous variables). Correlations between categorical variables were computed using Goodman and Kruskal's γ coefficient. The rates of rehospitalization and mortality were analyzed with Cox proportional hazard models, Kaplan-Meier curves, and Log-rank tests. Cox proportional hazard models were constructed to elucidate independent prognostic values of region information after adjustment for ACC/AHA HF stage, NYHA functional class, and the MAGGIC risk score. Furthermore, to evaluate the improvement of Cox proportional hazard models by adding region information to these risk predictors, integrated discrimination improvement, net reclassification improvement, and median improvement in risk score were calculated using R package survIDINRI (version 1.1-1).

A 2-sided p-value of <0.05 was considered statistically significant. We used R (version 3.4.0, R Foundation for Statistical Computing, Vienna, Austria) for all statistical analyses (165).

3.2 Study population and methods of machine learning-based mortality prediction among patients undergoing CRT implantation

3.2.1 Study population and protocol

We identified 2,282 patients who underwent successful CRT implantation at the Heart and Vascular Center of Semmelweis University (Budapest, Hungary) between September 2000 and December 2017. For each patient, pre-implantation clinical characteristics such as demographics, medical history, physical status, and vitals, currently applied medical therapy, ECG, echocardiographic, and laboratory parameters were extracted retrospectively from electronic medical records and entered into our structured database.

An additional prospective database of patients undergoing CRT implantation between January 2009 and December 2011 was also utilized. Patients included in both the retrospective and the prospective databases were removed from the retrospective database. In this way, the two cohorts were completely independent, and they could be used as training and test cohorts for ML algorithms.

The study protocol (Figure 9) complies with the Declaration of Helsinki, and it was approved by the Regional and Institutional Committee of Science and Research Ethics (approval No. 161/2019).

3.2.2 Feature selection and data pre-processing

Our structured database initially comprised over 100 easily obtainable pre-implantation clinical variables. Firstly, features included in both the retrospective and the prospective databases were identified (n = 49). Then, features missing for >40% of cases (n = 16) were excluded. The final set of input features included 33 pre-implantation clinical variables (Table 1).

Missing values were imputed using the mean imputation method, which replaces the missing values of a certain variable with the mean of the available cases. As the range of different features varied widely and some of the utilized algorithms required the data to be normalized, Z-score transformation was performed after imputation.

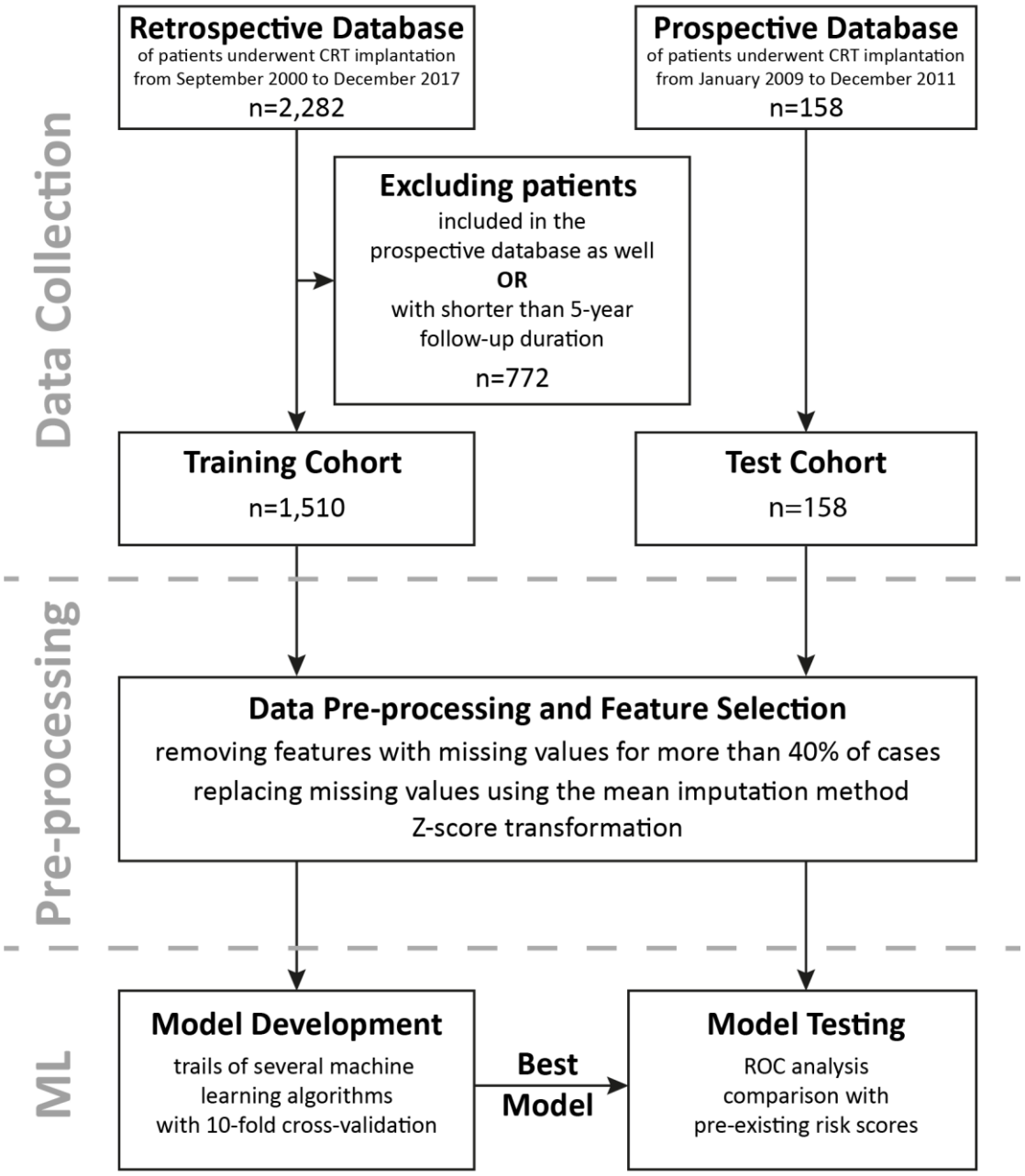


Figure 9 *Outline of the study protocol (166)*

CRT – cardiac resynchronization therapy, ML – machine learning, ROC – receiver operating characteristic

Table 1 *Pre-implantation clinical features included in the analysis*

Demographics and clinical characteristics	Comorbidities	ECG	Laboratory results	Medications
age at CRT implantation	hypertension	QRS duration	hemoglobin	ACE-I/ARB
gender	diabetes mellitus	QRS morphology	lymphocyte	beta-blockers
height	atrial fibrillation		serum sodium	furosemide
weight			total cholesterol	other loop diuretics
systolic blood pressure			serum creatinine	thiazide diuretics
NYHA functional class			GFR	MRA
etiology of heart failure			serum uric acid	statin
LVEF			serum urea	amiodarone
type of implanted device			NT-proBNP	digitalis
				allopurinol

ACE-I – angiotensin-converting enzyme inhibitors, ARB - angiotensin II receptor blockers, CRT – cardiac resynchronization therapy, GFR – glomerular filtration rate, LVEF – left ventricular ejection fraction, MRA – mineralocorticoid receptor antagonists, NT-proBNP – N-terminal pro-brain natriuretic peptide, NYHA – New York Heart Association

3.2.3 Model development

We used the follow-up data to generate six classes of possible outcomes: death during the 1st (class 1), the 2nd (class 2), the 3rd (class 3), the 4th (class 4), the 5th year after CRT implantation (class 5), and no death during the first 5 years following the implantation (class 6). The task of ML algorithms was to predict the probability distribution (i.e., class membership probabilities) of each patient over these classes based on the pre-implantation clinical features.

Model development included trials of several ML classifiers such as logistic regression, ridge regression, SVM, KNN classifier, gradient boosting classifier, traditional random forest (TRF), conditional inference random forest (CIRF), and multi-layer perceptron. Models were trained with stratified 10-fold cross-validation on the training cohort, and a grid search approach was used to tune the hyper-parameters of each ML algorithm.

The outputs of each model were series of six values representing the previously defined class membership probabilities (Figure 10A). The sum of these probabilities is equal to one in each patient. To create binary classifiers, we calculated cumulative class membership probabilities by summing these values until the given year of follow-up (Figure 10B). The computed cumulative probabilities were then calibrated using Platt's scaling, and the survival curve could be plotted for each patient (Figure 10C). The calibration of the model was evaluated using the Brier score (ranging from 0 to 1, with 0 representing the best possible calibration), which is defined as the mean squared

difference between the observed and the predicted outcome. Expected survival was also calculated from the annual calibrated cumulative probabilities (Figure 10D).

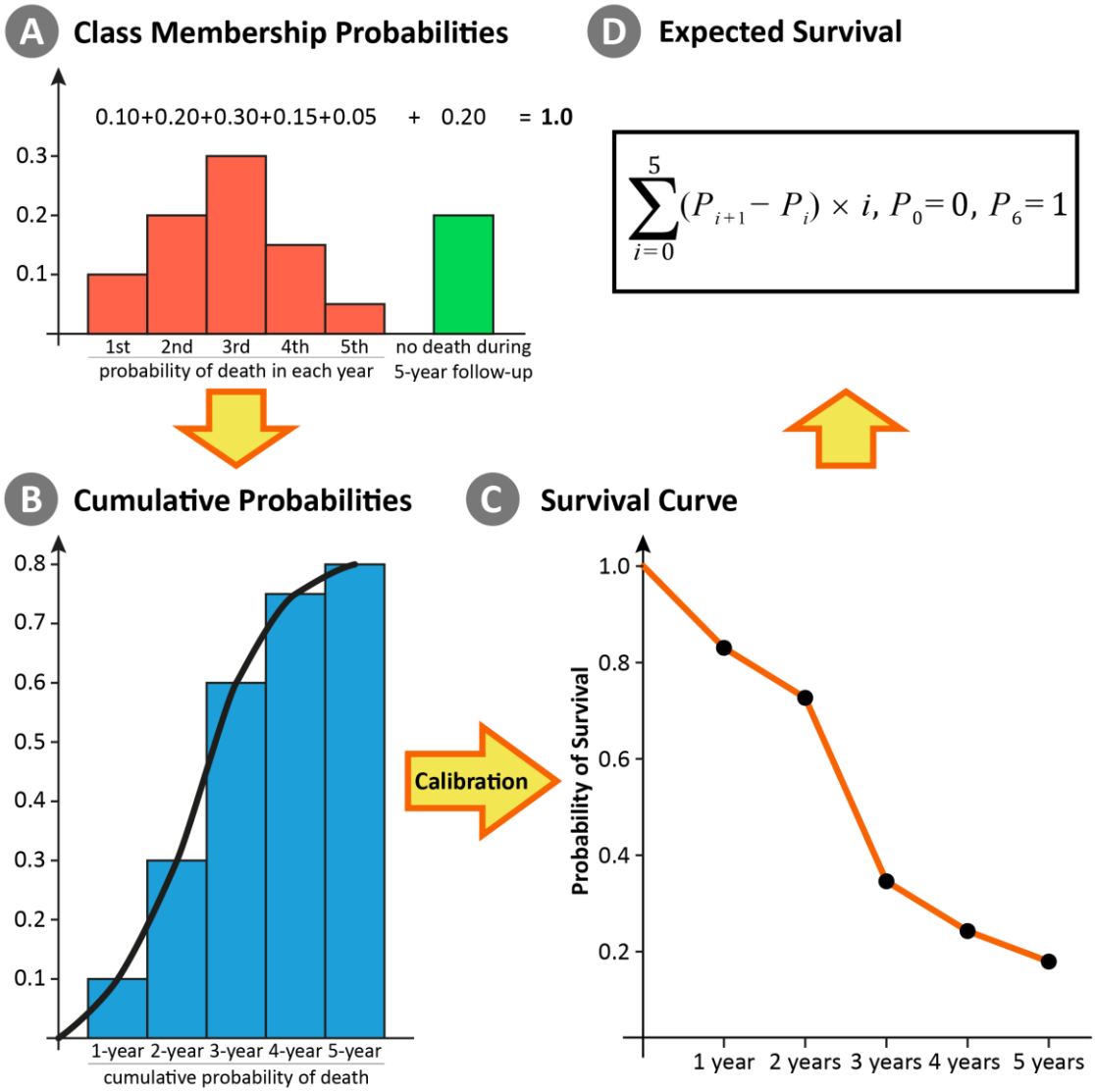


Figure 10 *Computation of survival probabilities in a representative patient (166)*

(A) The outputs of each model were series of class membership probabilities. (B) Cumulative probabilities were calculated by summing these values until the given year of follow-up. (C) To calibrate the cumulative probabilities, Platt's scaling was performed. Using these calibrated cumulative probabilities, the survival curve could be plotted for each patient. (D) Then, the expected survival time of each patient was estimated from the annual survival probabilities.

P_i – the calibrated cumulative probability of all-cause mortality at year i

To quantify the model's discriminative capabilities in each year, ROC curve analysis was performed, and AUC was calculated. The mean AUC of 1-, 2-, 3-, 4-, and 5-year calibrated cumulative probabilities was computed, and it served as the major metric to assess a model's performance.

3.2.4 Model testing

The model with the highest mean AUC was selected for further evaluation, and it is referenced as the SEMMELWEIS-CRT (perSonalizEd assessMent of estiMatEd risk of mortaLity With machinE learnIng in patientS undergoing CRT implantation) score throughout the entire thesis. To determine whether the model remains accurate when new data are fed into it, we tested it on the patients of the test cohort.

For each patient in the test cohort, we also computed pre-existing risk scores (SHFM, VALID-CRT, EAARN, ScREEN, and CRT-score) (167-171). Their prediction capabilities were quantified annually with AUCs, and they were compared with SEMMELWEIS-CRT score using the DeLong test.

3.2.5 Feature importances

To determine the major predictors of all-cause mortality in our patient population, permutation feature importances were computed from the final model. Permutation feature importance measures the importance of an input feature by calculating the increase in the model's prediction error after permuting its values. A feature is considered important if shuffling its values decreases the model's discriminative capability as the model relies heavily on that feature for the prediction. A feature is unimportant if shuffling its values leaves the AUC unchanged, as in this case, the model ignores the feature for the prediction.

3.2.6 Statistical analysis

Continuous variables are expressed as median (interquartile range), while categorical variables are reported as frequencies and percentages. The baseline clinical characteristics of patients alive and dead were compared with unpaired Student's t-test or Mann-Whitney U test for continuous variables and Chi-squared or Fisher's exact test for categorical variables, as appropriate. The survival of subgroups was visualized on Kaplan-Meier curves, and Log-rank tests were performed for comparison. Cox proportional hazards models were used to compute hazard rations (HRs) with 95% confidence intervals (CIs). A 2-sided p-value of <0.05 was considered statistically significant.

3.2.7 Software packages

ML algorithms were implemented in Python (version 3.6.7, Python Software Foundation, Wilmington, Delaware, U.S.A.) using the scikit-learn library (version 0.20.2) (172). Statistical analysis, including group comparisons and survival analysis, was performed in R (version 3.4.2, R Foundation for Statistical Computing, Vienna, Austria) (165).

3.3 Study population and methods of exploring sex-specific patterns of mortality predictors among patients undergoing CRT implantation

3.3.1 Study population and protocol

We identified 2,412 patients with chronic HFrEF (NYHA functional class II-IV) who underwent successful CRT implantation at the Heart and Vascular Center of Semmelweis University (Budapest, Hungary) between September 2000 and September 2018. For each patient, pre-implantation clinical characteristics (demographics, medical history, physical status, vitals, currently applied medical therapy, ECG, echocardiographic and laboratory parameters) and procedural parameters (type of the implanted device, LV lead position) were collected retrospectively from paper-based or electronic medical records and entered to our structured database.

The study protocol complies with the Declaration of Helsinki, and it was approved by the Regional and Institutional Committee of Science and Research Ethics (Approval No. 161/2019).

3.3.2 Study outcomes

Follow-up data (status [dead or alive], date of death) was obtained for all patients by querying the National Health Insurance Database of Hungary in September 2019. Accordingly, all patients included in our database were followed for at least 1 year or died within 1 year. In the entire study population, 2,116 patients also had 3-year outcome data available. The primary endpoint of our study was all-cause mortality.

3.3.3 Feature selection and data pre-processing

The data analysis pipeline, including feature selection, data pre-processing, ML model development and evaluation, is illustrated in Figure 11.

Feature selection included two consecutive steps. First, any feature with ≥40% missing data was removed. Second, collinear variables (Spearman correlation coefficient ≥ 0.3 or ≤-0.3) were also excluded as variables containing redundant information might bias the further steps of the analysis. The excluded features due to collinearity were serum urea (r = 0.748 vs. serum creatinine), glomerular filtration rate (r = -0.849 vs. serum creatinine), history of myocardial infarction (r = 0.803 vs. etiology of HF), history of percutaneous coronary intervention (r = 0.675 vs. etiology of HF), history of coronary artery bypass grafting (r = 0.383 vs. etiology of HF), history of ventricular arrhythmia (r = 0.313 vs. amiodarone), and oral anticoagulants (r = 0.364 vs. type of atrial fibrillation [AF]). As body weight and height also correlated moderately (r = 0.582), we used BMI instead. The final set of input features comprised 30 pre-implantation and procedural variables: baseline demographics and clinical characteristics (n = 10), comorbidities (n = 6), ECG- (n = 1), laboratory parameters (n = 3), and currently applied medications (n = 10). The list of candidate variables and the feature selection procedure are presented in Table 2.

Patients with more than 30% of missing values were excluded from further analyses. Missing values of continuous variables were imputed using MICE. This method models each feature with missing values as a function of other features in a round-robin fashion. As opposed to single imputations methods (e.g., mean imputation), it involves multiple imputations in order to mitigate the statistical uncertainty associated with imputations. Missing values of categorical variables were replaced with -1. As the range of different continuous features varied widely, Z-score transformation was applied after imputation to eliminate the possibility of model bias caused by the differing magnitude of the numerical values.

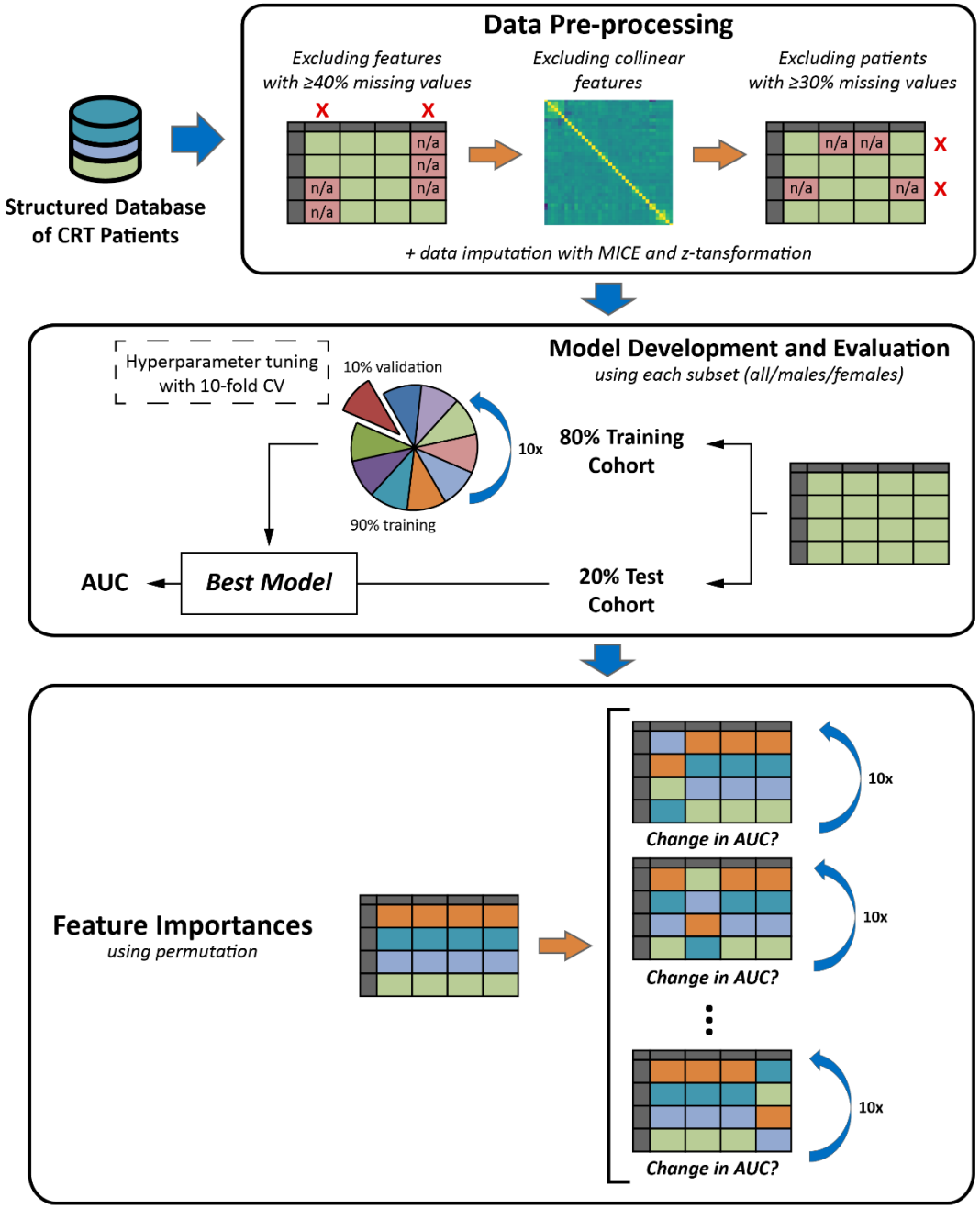


Figure 11 *The schematic outline of the data analysis pipeline (173)*

The data analysis pipeline included three major steps: (1) data pre-processing, (2) machine learning model development and evaluation, and (3) the calculation of feature importances. During data pre-processing, feature selection was performed, patients with a high proportion of missing data were excluded, missing values were imputed using MICE, and z-transformation was performed. Then, machine learning models were implemented in different subsets of patients. Before model training, each patient subset was split into training and test cohorts (80:20 ratio). Hyperparameter tuning was performed with 10-fold CV in each training cohort. Models' discriminatory power was estimated using the AUC. Each model was retrained in the given training cohort, and its performance was evaluated in the corresponding test cohort. Finally, to identify the most important predictors of mortality in each subset, permutation feature importances were computed from each of the final models.

AUC – area under the receiver operating characteristic curve, CRT – cardiac resynchronization therapy, CV – cross-validation, MICE – multiple imputation by chained equations.

Table 2 *Steps of feature selection and the list of clinical features included in the final models*

	Demographics and clinical characteristics	**Comorbidities**	**ECG**	**Laboratory parameters**	**Medications**
Included in the ML models	age at CRT implantation sex body mass index NYHA functional class HF duration >18 months etiology of heart failure left ventricular EF LV end-diastolic diameter type of implanted device LV lead position	hypertension diabetes mellitus type of AF COPD smoking status valvular heart disease	QRS morphology	hemoglobin serum sodium serum creatinine	ACE-I/ARB beta-blockers CCB loop diuretics thiazide diuretics MRA digitalis amiodarone statin allopurinol
Excluded due to collinearity	height weight	history of MI history of CABG and/or PCI		serum urea GFR	oral anticoagulants
Excluded due to ≥40% missing values	systolic blood pressure diastolic blood pressure heart rate LV end-diastolic volume LV end-systolic volume		QRS duration PR interval	lymphocyte total cholesterol serum uric acid NT-proBNP	

Feature selection included two consecutive steps. Firstly, features missing for more than 40% of patients were excluded. Then, collinear variables (Spearman correlation coefficient ≥0.3 or ≤-0.3) were also eliminated as highly correlated variables might bias the further steps of the analysis. The final set of features included 30 clinical variables: age at CRT implantation, sex, body mass index, New York Heart Association functional class, heart failure duration >18 months, etiology of heart failure (ischemic or non-ischemic), left ventricular ejection fraction and end-diastolic diameter assessed with two-dimensional echocardiography, type of the implanted device (CRT-P or CRT-D), left ventricular lead position (anterior, lateral or posterior), hypertension, diabetes mellitus, type of atrial fibrillation (paroxysmal, persistent, permanent), chronic obstructive pulmonary disease, smoking status, valvular heart disease (moderate to severe aortic valve disease, moderate to severe mitral valve disease, severe tricuspid regurgitation), QRS morphology (non-LBBB or LBBB), hemoglobin concentration, serum sodium and creatinine, medical treatment with angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers, beta-blockers, calcium channel blockers, loop diuretics, thiazide diuretics, mineralocorticoid receptor antagonists, digitalis, amiodarone, statins and allopurinol.

ACE-I – angiotensin-converting enzyme inhibitors, AF – atrial fibrillation, ARB - angiotensin II receptor blockers, CABG – coronary artery bypass grafting, CCB – calcium channel blockers, COPD – chronic obstructive pulmonary disease, CRT – cardiac resynchronization therapy, ECG – electrocardiogram, EF – ejection fraction, GFR – glomerular filtration rate, HF – heart failure, LV – left ventricular, MI – myocardial infarction, ML – machine learning, NYHA – New York Heart Association, NT-proBNP – N-terminal pro-brain natriuretic peptide, MRA – mineralocorticoid receptor antagonists, PCI – percutaneous coronary intervention

3.3.4 Model development and evaluation

We developed ML models to predict two separate outcomes: (1) 1-year all-cause mortality, and (2) 3-year all-cause mortality in the entire cohort, in males and females separately (a total of 6 separate binary classification tasks). To quantify a model's discriminatory power, ROC curve analysis was performed, and the AUC was calculated. Model development included trials of several binary classifiers such as logistic regression, SVM, KNN, Gradient Boosting classifier, TRF, CIRF, and multi-layer perceptron.

As the first step of model derivation, 20% of the given patient subset (all patients, males or females) was randomly selected as the holdout (test cohort). This split was performed in a stratified manner to ensure that the original ratio of outcomes was preserved in the training and test cohorts. Hyperparameter tuning was performed with stratified 10-fold cross-validation in the remaining data (80%, training cohort). The algorithm (with fine-tuned hyperparameters) exhibiting the highest AUC was then retrained in the entire training cohort, and its performance was evaluated in the test cohort in a statistically independent way. Finally, calibration of the ML models was assessed in the test cohort using Brier score (ranging from 0 to 1, with 0 representing the best possible calibration), which is defined as the mean squared difference between the observed outcomes and the predicted probabilities.

3.3.5 Feature importances

To determine the major predictors of 1- and 3-year all-cause mortality in each patient subset, permutation feature importances were computed from each of the 6 final models. Briefly, the importance of an input feature is measured by calculating the increase in the model's prediction error after permuting its values while keeping other features the same as before. In the current study, permutation was performed 10 times for each feature. A feature is considered important if shuffling its values decreases the model's discriminatory power (AUC) as the model relies heavily on that feature for the prediction. On the other hand, a feature is unimportant if shuffling its values leaves the AUC unchanged because, in this case, the model ignores the feature while predicting the outcome. After calculating the importance of each feature, we divided it by the AUC

measured in the data set before shuffling any of its features to enable the comparison of feature importances between different models.

3.3.6 Statistical analysis

Continuous variables are expressed as median (interquartile range), while categorical variables are reported as frequencies and percentages. The baseline clinical characteristics of patient subsets were compared with unpaired Student's t-test or Mann-Whitney U test for continuous variables and Chi-squared or Fisher's exact test for categorical variables, as appropriate. The survival of subgroups is presented using Kaplan-Meier curves. Cox proportional hazards models were used to compute HRs with 95% CIs. A 2-sided p-value of <0.05 was considered as statistically significant.

3.3.7 Software packages

Data pre-processing, feature selection, and ML algorithms were implemented in Python (version 3.6.8, Python Software Foundation, Wilmington, Delaware, U.S.A.) using the scikit-learn module (versions 0.21.3 and 0.23.dev0) (172). Statistical analyses, including group comparisons and survival analyses, were performed in R (version 3.6.1, R Foundation for Statistical Computing, Vienna, Austria) (165).

4. RESULTS

4.1 Applying topological data analysis to integrate echocardiographic features of left ventricular structure and function into a patient similarity network

4.1.1 The continuum of cardiac function

The use of TDA to create a patient-patient similarity network in the retrospective data set resulted in the formation of a looped structure. After the addition of cases from the prospective data set, the loop was persistent, which validated that the loop structure of the network model was intrinsic to the data and not an artifact. The combined network was used for discovering feature distributions and developing associations with clinical and outcome data. We noted that echocardiographic variables followed a gradually changing pattern throughout the loop (Figure 12). Moving counterclockwise, starting from the top of the loop, we observed gradually decreasing EF and e' values and increasing LV mass index, E/e' ratio, left atrial volume index, and TRV values. In addition to the echocardiographic features, cardiovascular risk factors, MACE-related rehospitalization, and death also showed accumulation in distinct sections of the loop (Table 3).

Upon seeing the gradations of echocardiographic variables, and to create clinically useful categories, we measured multi-dimensional Euclidean distance of nodes using 9 echocardiographic parameters that were used to create the network. The distance was used to create the groups – which was correlated with the gradation of the variables – on the network and subsequently collated based on clinical characteristics into 4 regions (Figure 13). Patients in each region showed differences in clinical characteristics (Figure 13, Tables 3 and 4). Notably, with progressing from the first to the fourth region, an increasing trend was seen in age and the prevalence of underlying risk factors and comorbidities. LVEF remained preserved in the first to third regions; however, it was significantly reduced in the fourth region ($p < 0.001$). LV mass progressively increased, and diastolic function parameters progressively worsened from region I to IV. We found a correlation between the regions and both NYHA functional classes ($\gamma = 0.47$; $p < 0.001$) and ACC/AHA stages ($\gamma = 0.52$; $p < 0.001$) in the prospective cohort, which showed more symptomatic patients in the third and fourth regions than in the first region (Figure 14A).

Table 3 *Clinical and echocardiographic characteristics of the retrospective cohort*

	Retrospective cohort (n = 866)	Region I (n = 177)	Region II (n = 138)	Region III (n = 286)	Region IV (n = 212)	p-value
Demographics and clinical characteristics						
Male, n (%)	387 (45)	73 (41)	39 (28)‡	102 (36)‡	142 (67)‡	<0.001
Age, years	66 (54-79)	49 (37-59)‡	64 (58-72)‡	78 (67-86)‡	66 (55-80)	<0.001
BMI, kg/m^2	26.8 (23.6-30.4)	25.9 (23.0-29.8)	25.7 (22.5-29.4)	27.0 (23.8-30.3)	27.5 (24.3-31.4)	0.013
Hypertension	441 (51)	48 (27)‡	54 (39)†	189 (66)‡	126 (59)†	<0.001
Hyperlipidemia	358 (41)	37 (21)‡	45 (33)*	153 (53)‡	99 (47)	<0.001
Diabetes	178 (21)	16 (9)‡	14 (10)†	79 (28)‡	60 (28)†	<0.001
COPD	61 (7)	4 (2)†	14 (10)	25 (9)	15 (7)	0.012
Tobacco abuse	377 (44)	76 (43)	50 (36)	122 (43)	105 (50)	0.282
History of CKD	214 (25)	18 (10)‡	18 (13)‡	96 (34)‡	62 (29)*	<0.001
Clinical outcomes						
MACE rehosp.	147 (17)	8 (5)‡	10 (7)‡	77 (27)‡	45 (21)	<0.001
MACE death	10 (1)	0 (0)	0 (0)	3 (1)	5 (2)*	0.083
Echocardiography						
LVEF, %	62 (57-67)	63 (59-66)†	65 (61-68)‡	64 (60-68)‡	56 (45-74)‡	<0.001
LVMi, g/m^2	85 (67-106)	66 (58-77)‡	64 (58-76)‡	93 (78-115)‡	105 (87-129)‡	<0.001
E, m/s	0.79 (0.60-0.90)	0.73 (0.63-0.88)†	0.70 (0.60-0.80)‡	0.80 (0.70-1.00)‡	0.80 (0.60-1.00)†	<0.001
A, m/s	0.76 (0.60-1.00)	0.58 (0.50-0.67)‡	0.86 (0.72-0.93)‡	1.11 (0.91-1.25)‡	0.61 (0.49-0.75)‡	<0.001
E/A	0.90 (0.80-1.20)	1.30 (1.10-1.50)‡	0.80 (0.70-0.90)‡	0.80 (0.70-0.90)‡	1.25 (1.00-1.90)‡	<0.001
e', cm/s	6.0 (4.4-7.6)	9.0 (8.0-11.0)‡	6.7 (6.0-7.0)‡	4.3 (4.0-5.0)‡	5.7 (4.5-7.0)†	<0.001
E/ e'	12.5 (9.2-17.9)	8.3 (7.1-9.9)‡	10.4 (8.5-12.0)‡	18.2 (14.5-24.5)‡	13.8 (10.4-19.5)†	<0.001
LAVi, mL/m^2	34 (27-43)	28 (22-34)‡	27 (22-33)‡	38 (31-48)‡	42 (33-55)‡	<0.001
TRV, m/s	2.30 (2.00-2.70)	2.10 (1.90-2.30)‡	2.20 (2.00-2.50)*	2.50 (2.20-2.80)‡	2.40 (2.10-2.84)†	<0.001

Values are n (%) or median (interquartile range) for continuous variables.

*‡p < 0.001; †p < 0.01; *p < 0.05 between the given region and the remaining regions, Kolmogorov-Smirnov test. Overall p-values are calculated using analysis of variance (ANOVA) or Kruskal-Wallis test.*

A – late diastolic transmitral flow velocity, BMI – body mass index, CKD – chronic kidney disease, COPD – chronic obstructive pulmonary disease, E – early diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, LAVi – left atrial volume index, LVEF – left ventricular ejection fraction, LVMi – left ventricular mass index, MACE – major adverse cardiovascular event, TRV – tricuspid regurgitation peak velocity

Table 4 *Clinical and echocardiographic characteristics of the prospective cohort*

	Prospective cohort (n = 468)	**Region I (n = 165)**	**Region II (n = 113)**	**Region III (n = 59)**	**Region IV (n = 112)**	**p-value**
Demographics and clinical characteristics						
Male, n (%)	195 (42)	72 (44)	33 (29)†	12 (20)‡	68 (61)‡	<0.001
Age, years	57 (47-66)	46 (36-56)‡	60 (53-66)‡	66 (58-75)‡	63 (53-71)†	<0.001
BMI, kg/m2	30.7 (25.7-36.5)	29.5 (25.1-37.1)	31.5 (25.1-37.8)	28.6 (24.4-36.3)	31.0 (27.3-34.9)	0.509
SBP, mmHg	126 (114-140)	122 (110-135)*	124 (115-140)	135 (122-147)†	127 (111-147)	<0.001
DBP, mmHg	75 (68-82)	75 (69-83)	76 (68-81)	72 (68-80)	73 (66-79)	0.252
Hypertension	318 (68)	88 (53)‡	79 (70)	45 (76)	91 (81)‡	<0.001
Hyperlipidemia	337 (72)	101 (61)‡	89 (79)	47 (80)	84 (75)	0.003
Diabetes	115 (25)	22 (13)‡	26 (23)	16 (27)	43 (38)‡	<0.001
COPD	53 (19)	13 (8)	7 (6)*	4 (7)	28 (25)‡	<0.001
Tobacco abuse	206 (44)	63 (38)*	53 (47)	26 (44)	58 (52)	0.131
History of CAD	114 (24)	12 (7)‡	22 (19)	19 (32)	55 (49)‡	<0.001
History of CVA[#]	31 (11)	7 (7)	4 (6)	3 (12)	15 (21)†	0.026
History of CKD	81 (17)	5 (3)‡	12 (11)	16 (27)*	41 (37)‡	<0.001
HF diagnosed ≥ 18 months	68 (14)	1 (1)‡	2 (2)‡	10 (17)	53 (47)‡	<0.001
Clinical outcomes						
NYHA class, I/II/III/IV	268/123/62/15	123/32/9/1‡	63/44/6/0‡	25/25/9/0*	44/19/35/14‡	<0.001
ACC/AHA stage, A/B/C/D	122/168/171/7	80/43/42/0‡	26/37/50/0	4/28/26/1†	8/48/50/6‡	<0.001
MAGGIC score	11 (7-16)	9 (5-12)‡	12 (9-15)†	14 (11-20)‡	18 (9-26)‡	<0.001
MACE rehosp.	147 (17)	8 (5)‡	10 (7)‡	77 (27)‡	45 (21)	<0.001
MACE death	10 (1)	0 (0)	0 (0)	3 (1)	5 (2)*	0.083
Echocardiography						
LVEF, %	62 (57-66)	62 (59-65)†	65 (62-68)‡	63 (59-67)	54 (39-60)‡	<0.001
LVMi, g/m^2	74 (60-95)	64 (55-75)‡	64 (56-76)‡	83 (71-102)†	108 (89-136)‡	<0.001
E, m/s	0.79 (0.65-0.94)	0.81 (0.69-0.92)*	0.69 (0.60-0.82)‡	0.81 (0.63-0.94)	0.90 (0.69-1.11)‡	<0.001
A, m/s	0.71 (0.57-0.87)	0.62 (0.52-0.72)‡	0.86 (0.75-0.94)‡	0.99 (0.85-1.16)‡	0.58 (0.42-0.70)‡	<0.001
E/A	1.10 (0.82-1.40)	1.28 (1.10-1.50)‡	0.82 (0.73-0.91)‡	0.79 (0.70-0.93)‡	1.43 (1.10-2.10)‡	<0.001
e', cm/s	7.0 (5.3-9.0)	9.1 (8.0-11.0)‡	7.0 (6.0-8.0)‡	5.0 (4.0-5.6)‡	5.0 (4.0-6.5)‡	<0.001
E/ e'	10.5 (8.4-14.4)	8.4 (6.9-10.1)‡	10.2 (9.0-11.9)‡	15.5 (13.4-19.4)‡	16.5 (11.2-25.0)‡	<0.001
LAVi, mL/m^2	29 (22-37)	25 (21-31)‡	24 (19-29)‡	36 (28-43)‡	42 (34-56)‡	<0.001
TRV, m/s	2.19 (1.90-2.50)	2.01 (1.70-2.30)‡	2.10 (1.81-2.35)	2.40 (2.10-2.63)*	2.42 (2.11-2.90)‡	<0.001

Values are n (%) or median (interquartile range) for continuous variables.

*‡p < 0.001; †p < 0.01; *p < 0.05 between the given region and the remaining regions, Kolmogorov-Smirnov test. Overall p-values are calculated using analysis of variance (ANOVA) or Kruskal-Wallis test.*

[#]data available only for 274 patients

A – late diastolic transmitral flow velocity, ACC/AHA – American College of Cardiology / American Heart Association heart failure classification, BMI – body mass index, CAD – coronary artery disease, CKD – chronic kidney disease, COPD – chronic obstructive pulmonary disease, CVA – cerebrovascular accident, DBP – diastolic blood pressure, E – early diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, HF – heart failure, MACE – major adverse cardiovascular event, LAVi – left atrial volume index, LVEF – left ventricular ejection fraction, LVMi – left ventricular mass index, MACE – major adverse cardiovascular event, MAGGIC – Meta-Analysis Global Group in Chronic Heart Failure, NYHA – New York Heart Association functional classification of heart failure, SBP – systolic blood pressure, TRV – tricuspid regurgitation peak velocity

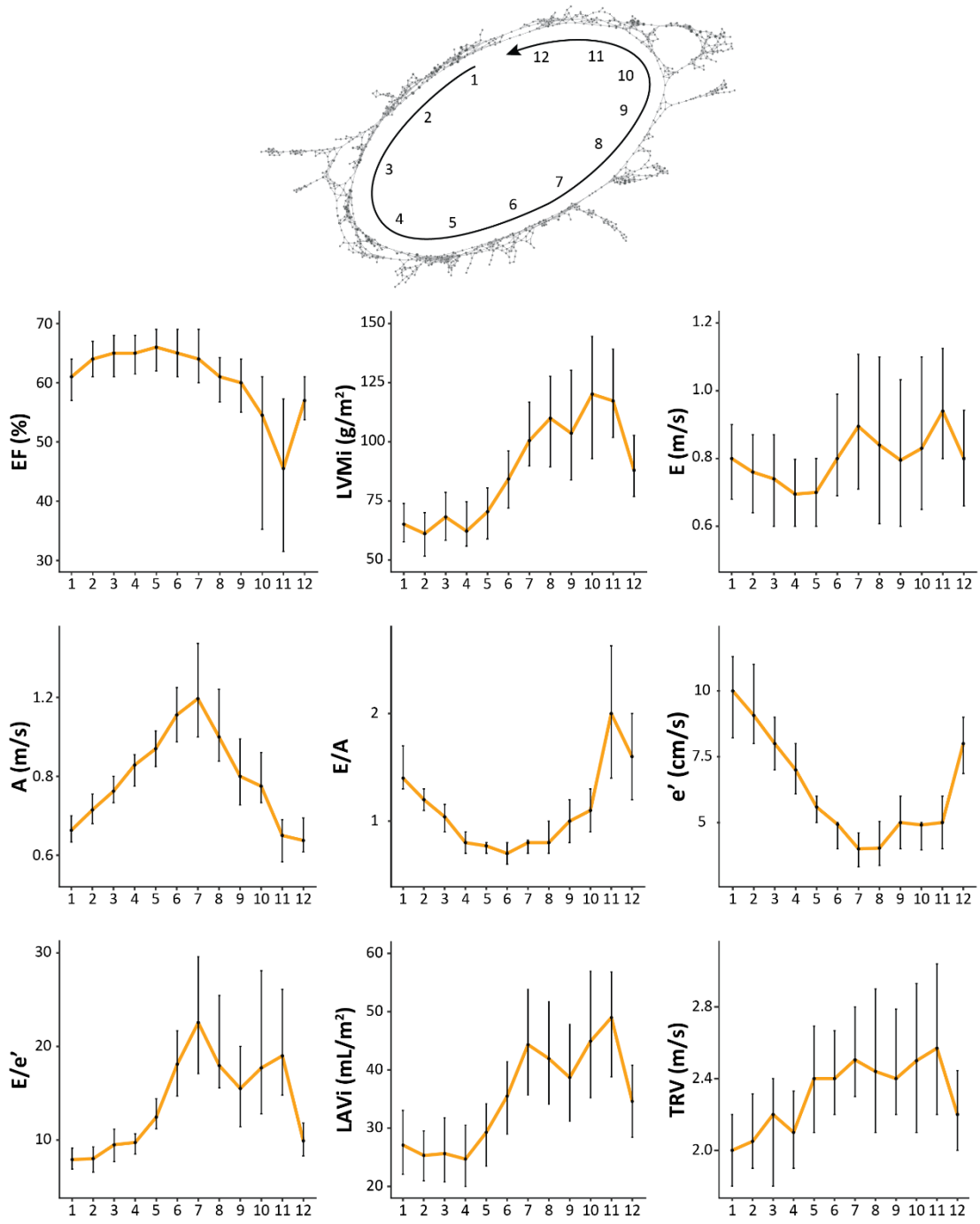


Figure 12 *Gradation of echocardiographic features on the topological network (154)*

The echocardiographic variables follow a gradually changing pattern while moving counterclockwise on the loop.

A – late diastolic transmitral flow velocity, E – early diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, LAVi – left atrial volume index, EF – ejection fraction, LVMi – left ventricular mass index, TRV – tricuspid regurgitation peak velocity

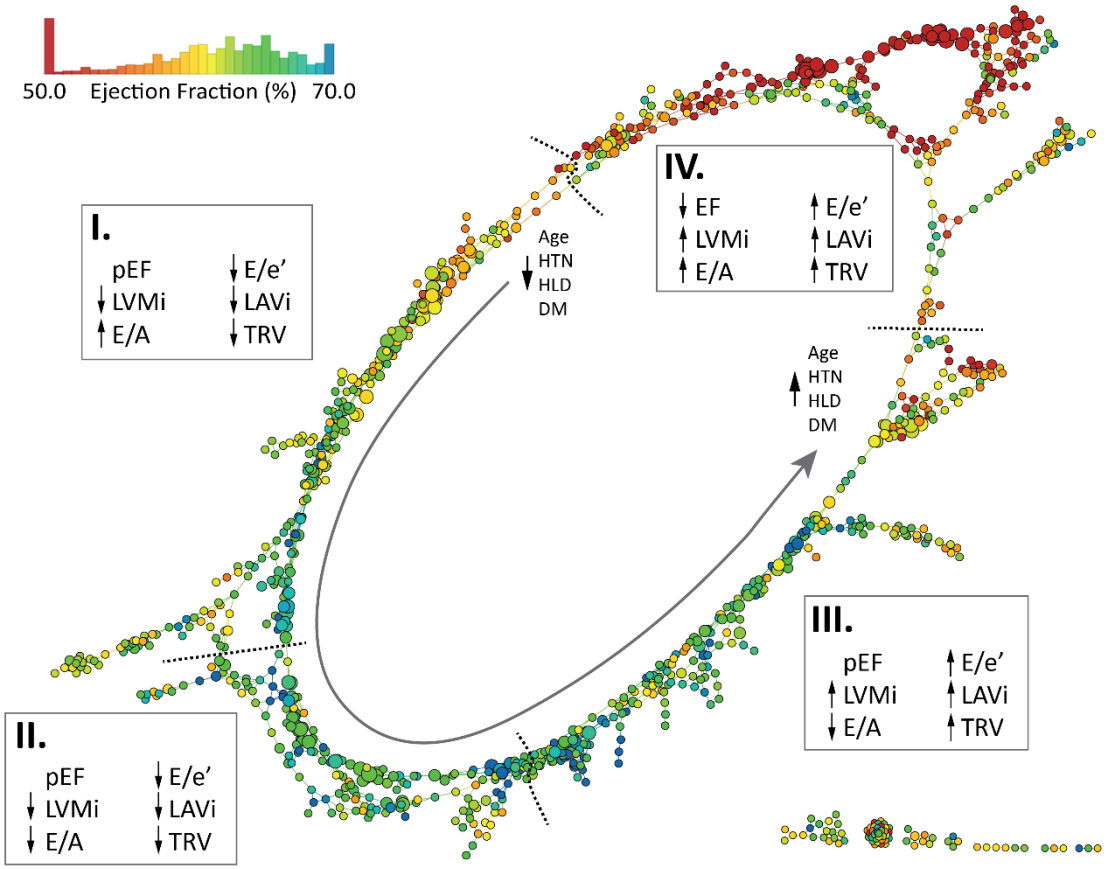


Figure 13 *The looped network of cardiac dysfunction (154)*

Multi-parametric echocardiographic data sets were used to develop a patient-patient similarity network using topological data analysis. Nodes indicate one or more patient(s) who have similar echocardiographic characteristics, and nodes having at least one patient in common are connected by edges. Nodes were color-coded according to the mean EF of the patients in the given node. Moving counterclockwise along the gray arrow, 4 regions were identified with different clinical presentations and outcomes. Region I mainly consisted of patients with risk factors but no obvious symptoms or disease. Patients in region II had more cardiac risk factors (especially HTN) with impaired LV relaxation. In region III, patients had advanced diastolic dysfunction, heart failure with pEF, and pulmonary HTN more frequently, whereas region IV included patients dominantly with reduced EF, increased LVMi, LAVi, and TRV. Although the map was developed using cross-sectional data, distinct regions of the networks correspond to distinct stages of the disease, along which patients can move on the map, signaling progression, treatment, and recovery of the disease.

A – late diastolic transmitral flow velocity, DM – diabetes mellitus, E – early diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, EF – ejection fraction, HLD – hyperlipidemia, HTN – hypertension, LAVi – left atrial volume index, pEF – preserved left ventricular ejection fraction, LVMi – left ventricular mass index, TRV – tricuspid regurgitation peak velocity

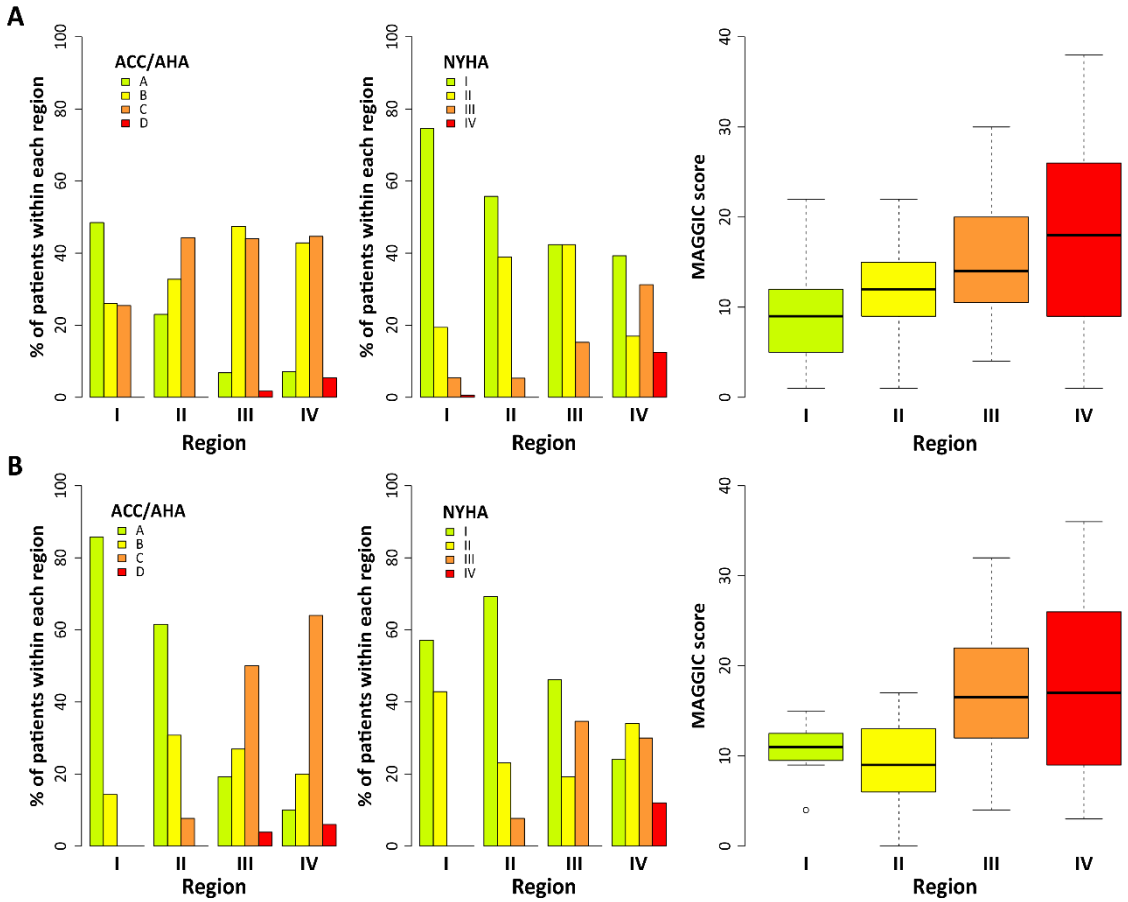


Figure 14 *Heart failure stages and classification by the four regions of the topological network (154)*

NYHA functional classes, ACC/AHA heart failure stages, and MAGGIC scores in the four regions of the (A) primary and (B) secondary cohort.

ACC – American College of Cardiology, AHA – American Heart Association. MAGGIC – Meta-Analysis Global Group in Chronic Heart Failure, NYHA – New York Heart Association

4.1.2 The association of regions with major adverse cardiac events

The median follow-up time in the primary cohort was 309 (100 – 531) days. A total of 207 (16%) patients were observed to have MACE-related hospitalizations, and 19 (1%) patients died due to MACE during follow-up. The number of MACE-related hospitalizations increased progressively from regions I to IV ($p < 0.001$), with MACE-related deaths seen only in the third and fourth regions ($p < 0.001$). The Kaplan-Meier curves for MACE-related rehospitalization in the regions differed significantly ($p < 0.001$) (Figure 15A). Patients in the fourth region had a >5-fold increased risk of re-hospitalization (HR: 5.89; 95% CI: 3.39 – 10.24; $p < 0.001$), whereas patients in the third region had a >6-fold increased risk of rehospitalization (HR: 6.88; 95% CI: 3.98 – 11.90; $p < 0.001$) compared with the first region. Subjects in the second region did not have a significantly increased risk of hospital admission due to MACE (HR: 1.45; 95% CI: 0.72 – 2.93; $p = 0.301$) compared with those in the first region. The number of MACE-related deaths was low in the first and second regions, whereas the probability of death was significantly higher in the third and fourth regions than in the first region ($p < 0.001$) (Figure 15B).

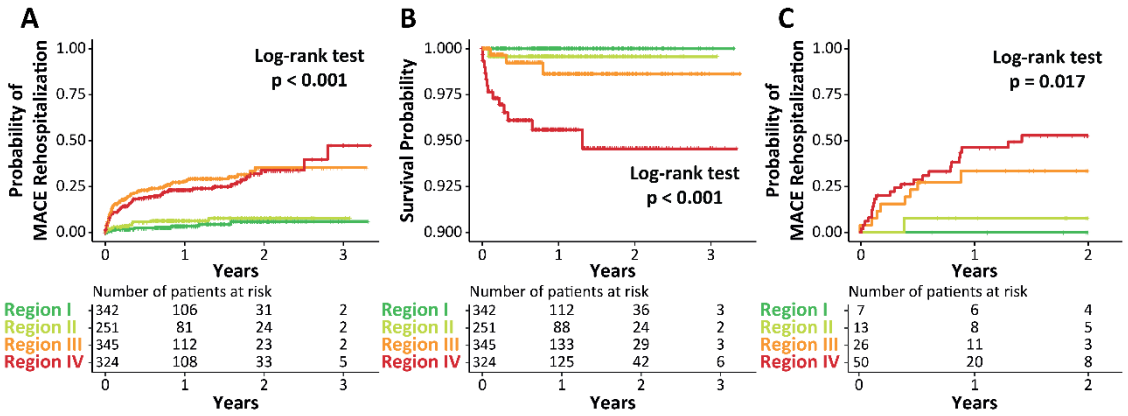


Figure 15 *Kaplan-Meier curves for MACE-related outcomes by regions (154)*

Kaplan-Meier curves of the four regions: (A) MACE-related rehospitalization in the primary cohort, (B) MACE-related death in the primary cohort, and (C) MACE-related rehospitalization in the secondary cohort.

MACE – major adverse cardiovascular event

4.1.3 Individualized patient predictions

Individualized patient predictions for clinical stages, severity, and future adverse events were tested in a secondary cohort analysis. Detailed demographics and between-region comparisons in this cohort are shown in Table 5. After predicting the region membership of patients of the secondary cohort using a random forest classifier, the same tendency was observed for the probability of MACE-related rehospitalization in the regions as in the primary cohort (Figure 15C). Subjects predicted to be in the fourth region had a >2-fold increased risk of MACE-related rehospitalization (HR: 2.75; 95% CI: 1.27 – 45.95; $p = 0.010$), whereas patients belonging to the third and second regions were not associated with a significantly increased risk compared with those in the first region. Patients in the first region were free of any MACE. A correlation between NYHA functional classes and ACC/AHA stages with regions was also observed ($\gamma = 0.56$ and $\gamma = 0.67$; both $p < 0.001$, respectively), which indicated that more symptomatic patients were found in the fourth region than in other regions (Figure 14B).

We also wanted to demonstrate whether changing the location of a patient on the loop was associated with worsening or improvement of cardiac function. To illustrate the motion of patients on the loop, the predicted regions of the first and second echocardiograms were compared (Figure 16). Both echocardiograms in 13 patients were in low-risk regions (region I or II), whereas those in 63 patients were in the high-risk region (region III or IV). Fifteen patients showed improvement (moved from region III and/or IV to region I and/or II) in echocardiographic results, and 5 patients showed worsening (moved from region I and/or II to region II and/or IV) echocardiographic results. Improvement or staying in the low-risk regions was associated with lower MACE-related rehospitalization rates following the second echocardiogram than worsening or staying in high-risk regions (3% vs. 37%, $p < 0.001$).

Table 5 *Clinical and echocardiographic characteristics of the longitudinal cohort*

	Longitud. cohort (n = 96)	Region I (n = 7)	Region II (n = 13)	Region III (n = 26)	Region IV (n = 50)	p-value
Demographics and clinical characteristics						
Male, n (%)	47 (49)	3 (43)	4 (31)	7 (27)*	33 (66)‡	0.004
Age, years	59 (50-67)	53 (30-59)*	55 (46-59)	65 (57-72)*	57 (49-66)	0.005
BMI, kg/m2	31.0 (27.4-34.7)	24.2 (22.2-27.9)†	29.3 (28.3-34.7)	31.4 (27.6-34.6)	31.6 (28.2-35.1)	0.020
SBP, mmHg	131 (117-147)	117 (111-134)	126 (120-142)	145 (123-151)	130 (112-149)	0.100
DBP, mmHg	79 (72-84)	77 (71-84)	82 (72-86)	77 (73-80)	80 (72-84)	0.684
Hypertension	83 (86)	3 (43)†	12 (92)	26 (100)*	42 (84)	0.002
Hyperlipidemia	75 (78)	4 (57)	9 (69)	23 (88)	39 (78)	0.219
Diabetes	38 (40)	0 (0)*	4 (31)	15 (58)*	19 (38)	0.029
COPD	27 (28)	3 (43)	2 (15)	7 (27)	15 (30)	0.620
Tobacco abuse	49 (51)	4 (57)	6 (46)	11 (42)	28 (56)	0.681
History of CAD	57 (59)	2 (29)	6 (46)	16 (62)	33 (66)	0.203
History of CVA	21 (22)	1 (14)	5 (38)	4 (15)	11 (22)	0.430
History of CKD	23 (24)	0 (0)	0 (0)*	8 (31)	15 (30)	0.034
HF diagnosed ≥18 months	37 (39)	0 (0)*	0 (0)†	10 (38)	27 (54)†	<0.001
Clinical outcomes						
NYHA class, I/II/III/IV	37/28/25/6	4/3/0/0	9/3/1/0	12/5/9/0	12/17/15/6†	0.026
ACC/AHA stage, A/B/C/D	24/22/46/4	6/1/0/0†	8/4/1/0†	5/7/13/1	5/10/32/3†	<0.001
MAGGIC score	15 (9-22)	11 (10-13)	9 (6-13)*	17 (12-22)	17 (9-26)	0.009
MACE rehosp.	32 (33)	0 (0)	1 (8)	8 (30)	23 (46)†	0.009
MACE death	4 (4)	0 (0)	0 (0)	0 (0)	4 (8)	0.525
Echocardiography						
LVEF, %	56 (47-62)	55 (54-59)	63 (60-69)†	59 (54-64)	50 (33-58)‡	<0.001
LVMi, g/m^2	96 (75-117)	54 (48-62)‡	60 (40-73)‡	94 (85-112)	103 (94-126)‡	<0.001
E, m/s	0.79 (0.67-1.01)	0.67 (0.62-0.87)	0.72 (0.67-0.79)	0.78 (0.69-1.08)	0.83 (0.69-1.08)	0.221
A, m/s	0.75 (0.54-0.94)	0.64 (0.48-0.65)*	0.89 (0.79-0.91)	1.02 (0.91-1.19)‡	0.60 (0.42-0.75)‡	<0.001
E/A	1.06 (0.80-1.67)	1.30 (1.26-1.52)	0.88 (0.79-0.92)*	0.77 (0.68-0.91)‡	1.64 (1.07-2.18)‡	<0.001
e', cm/s	6.0 (5.0-8.0)	10.0 (8.5-10.5)*	8.0 (6.0-8.0)	6.0 (4.0-7.0)	6.0 (4.0-8.0)	0.002
E/ e'	12.6 (9.5-17.1)	8.2 (6.2-10.7)*	9.8 (8.6-13.2)*	15.0 (11.6-18.0)	13.1 (9.9-21.2)	0.001
LAVi, mL/m^2	31 (22-41)	20 (16-22)†	22 (19-25)†	32 (28-36)	39 (28-48)‡	<0.001
TRV, m/s	2.33 (1.88-2.69)	1.86 (1.35-2.20)	2.10 (1.95-2.21)*	2.11 (1.76-2.69)	2.55 (2.06-2.94)†	0.016

Values are n (%) or median (interquartile range) for continuous variables.

*‡p < 0.001; †p < 0.01; *p < 0.05 between the given region and the remaining regions, Kolmogorov-Smirnov test. Overall p-values are calculated using analysis of variance (ANOVA) or Kruskal-Wallis test.*

A – late diastolic transmitral flow velocity, ACC/AHA – American College of Cardiology / American Heart Association heart failure classification, BMI – body mass index, CAD – coronary artery disease, CKD – chronic kidney disease, COPD – chronic obstructive pulmonary disease, CVA – cerebrovascular accident, DBP – diastolic blood pressure, E – early diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, HF – heart failure, MACE – major adverse cardiovascular event, LAVi – left atrial volume index, LVEF – left ventricular ejection fraction, LVMi – left ventricular mass index, MACE – major adverse cardiovascular event, MAGGIC – Meta-Analysis Global Group in Chronic Heart Failure, NYHA – New York Heart Association functional classification of heart failure, SBP – systolic blood pressure, TRV – tricuspid regurgitation peak velocity

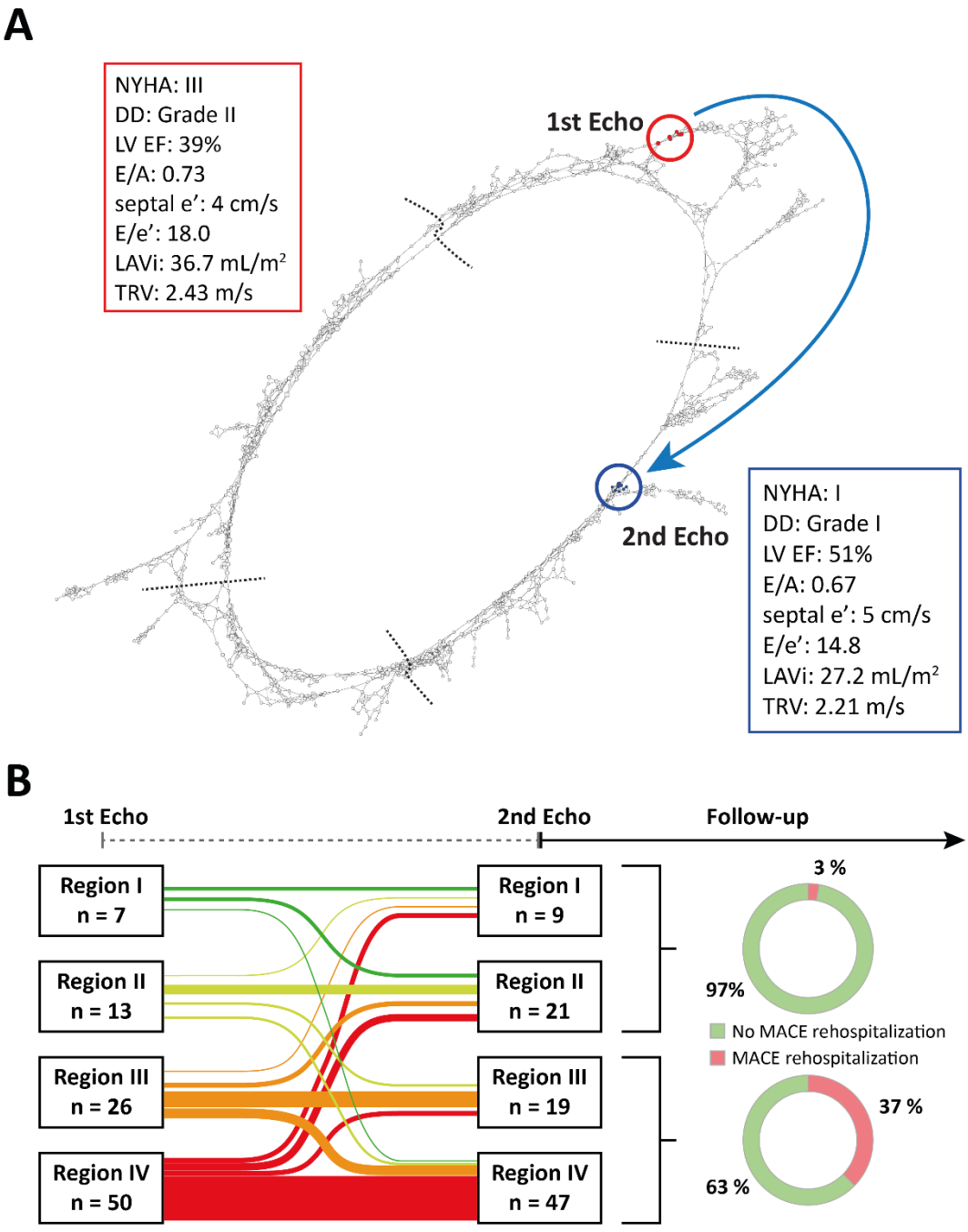


Figure 16 *Inter-region motion of secondary cohort patients on the topological network (154)*

(A) Pre- and post-operative position of a representative patient (74 years old, female) who underwent coronary artery bypass grafting. (B) Inter-region motion of patients in the secondary cohort.

A – late diastolic transmitral flow velocity, DD – diastolic dysfunction, E – early diastolic transmitral flow velocity, e' – early diastolic relaxation velocity at septal mitral annular position, LAVi – left atrial volume index, LVEF – left ventricular ejection fraction, MACE – major adverse cardiovascular event, NYHA – New York Heart Association functional class, TRV – tricuspid regurgitation peak velocity

4.1.4 Discrimination and reclassification

The incremental value of the topological regions was assessed in the prospective cohort. Even after adjustment for NYHA functional class, ACC/AHA HF stages, and MAGGIC scores, the predictive value of being in region IV remained significant with an 8- to 10-fold risk (Table 6). Net reclassification improvement, integrated discrimination improvement, and median improvement in risk score showed that adding region information to NYHA functional class, ACC/AHA stage, and MAGGIC score significantly improved the prediction of MACE-related events (Table 6). A combination of NYHA functional class (symptoms) and region information (cardiac function) performed better than that of the ACC/AHA stage, which also accounted for symptoms and cardiac function (AUC: 0.819 vs. 0.720).

Table 6 *Independency and incremental value of regions upon clinical risk predictors*

	Adjusted HR*			**Model improvement†**			
	HR	**95% CI**	**p-value**		**Estimate**	**95% CI**	**p-value**
Model with NYHA functional class + regions: AUC of 0.819 vs. 0.749 for model without regions							
Region II	2.20	0.53-9.21	0.280	IDI	0.079	0.017-0.166	0.007
Region III	3.90	0.92-16.46	0.064	NRI	0.775	0.077-0.899	0.027
Region IV	8.87	2.53-31.05	<0.001	MIRS	0.058	0.002-0.245	0.020
Model with ACC/AHA stage + regions: AUC of 0.815 vs. 0.720 for model without regions							
Region II	1.91	0.45-8.00	0.380	IDI	0.187	0.105-0.288	0.007
Region III	3.72	0.88-15.77	0.074	NRI	0.737	0.268-0.903	0.007
Region IV	12.39	3.71-41.35	<0.001	MIRS	0.275	0.118-0.386	0.013
Model with MAGGIC score + regions: AUC of 0.815 vs. 0.775 for model without regions							
Region II	2.01	0.48-8.43	0.340	IDI	0.098	0.033-0.192	0.007
Region III	2.92	0.67-12.76	0.160	NRI	0.437	0.071-0.821	0.020
Region IV	8.16	2,25-29.58	0.001	MIRS	0.146	0.025-0.250	0.013

**HR for each region adjusted for NYHA functional class, ACC/AHA stage, and MAGGIC score, respectively.*

†Model improvement by adding region information upon each risk predictor.

ACC/AHA – American College of Cardiology / American Heart Association heart failure classification, AUC – area under the receiver operating characteristic curve, CI – confidence interval, HR – hazard ratio, IDI – integrated discrimination improvement, MAGGIC – Meta-Analysis Global Group in Chronic Heart Failure, MIRS – median improvement in risk score, NRI – net reclassification improvement, NYHA – New York Heart Association functional classification

4.2 Designing and evaluating a machine learning-based risk stratification system to predict all-cause mortality of patients undergoing CRT implantation

4.2.1 Baseline clinical characteristics

The final training cohort included 1,510 patients (75.6% males, 45.6% CRT-D) who underwent CRT implantation. A total of 158 CRT patients (80.4% males, 20.3% CRT-D) were prospectively enrolled and entered into the test database. During the 5-year follow-up period, 805 (53.3%) patients died in the training cohort, and there were 80 (50.6%) deaths in the test cohort. Table 7 shows the baseline characteristics of both cohorts and the comparisons between patients who were dead and alive at 5-year follow-up.

4.2.2 Prediction of all-cause mortality

Among the evaluated ML classifiers, random forest (i.e., the SEMMELWEIS-CRT score) yielded the highest AUCs for predicting all-cause mortality at 1-, 2-, 3-, 4-, and 5-year follow-up in the test cohort (Figure 17 and Table 8). Calibration improved the Brier scores of the final model (Table 9). The leading predictors of all-cause mortality are presented in Figure 18.

4.2.3 Comparing the SEMMELWEIS-CRT score with pre-existing scores

When compared with the pre-existing risk scores, the SEMMELWEIS-CRT score demonstrated significantly better response prediction and greater discrimination of mortality (Table 8). The CRT-score exhibited the best performance among the pre-existing risk scores; however, our random forest-based classifier was still superior to it for the prediction of 5-year outcomes. Regarding the rest of the risk scores, the SEMMELWEIS-CRT score significantly outperformed them at all of the investigated time points.

Table 7 *Pre-implantation clinical characteristics of the training and test cohorts*

	Training cohort			Test cohort		
	All (n = 1,510)	**Alive (n = 705)**	**Dead (n = 805)**	**All (n = 158)**	**Alive (n = 78)**	**Dead (n = 80)**
Demographics and clinical characteristics						
Age, years	66 (60-74)	65 (58-71)	69 (62-76)‡	68 (61-74)	67 (60-71)	70 (63-75)*
Male	1,141 (75.6)	483 (68.5)	658 (81.7)‡	127 (80.4)	59 (75.6)	68 (85.0)
Weight, kg	80 (70-91)	80 (72-92)	80 (69-90)	80 (72-92)	80 (73-92)	80 (71-90)
Height, cm	172 (165-176)	171 (165-176)	172 (165-177)	172 (167-176)	172 (167-176)	172 (165-176)
SBP, mmHg	121 (110-134)	120 (111-136)	122 (110-133)	119 (109-134)	120 (110-135)	117 (108-134)
NYHA IV	274 (18.1)	137 (19.4)	137 (17.0)	15 (9.5)	9 (11.5)	6 (7.5)
CRT-D	688 (45.6)	310 (44.0)	378 (47.0)	32 (20.3)	14 (17.9)	18 (22.5)
QRS duration, ms	160 (141-180)	160 (140-172)	160 (148-190)	160 (150-180)	160 (160-180)	160 (140-180)
LBBB morphology	1,054 (69.8)	528 (74.9)	526 (65.3)‡	128 (81.0)	71 (91.0)	57 (71.3)†
Ischemic etiology	767 (50.8)	296 (42.0)	471 (58.5)‡	95 (60.1)	41 (52.6)	54 (67.5)
Atrial fibrillation	584 (38.7)	240 (34.0)	344 (42.7)‡	48 (30.4)	15 (19.2)	33 (41.3)†
Hypertension	1,055 (69.9)	480 (68.1)	575 (71.4)	97 (61.4)	43 (55.1)	54 (67.5)*
Diabetes mellitus	560 (37.1)	235 (33.3)	325 (40.4)†	64 (40.5)	26 (33.3)	38 (47.5)
Laboratory and echocardiographic parameters						
Hemoglobin, g/dL	13.6 (12.3-14.8)	13.9 (12.8-14.9)	13.3 (12.1-14.7)‡	14.0 (13.0-15.0)	14.0 (13.0-15.0)	13.9 (12.0-14.7)†
Lymphocyte, %	20.5 (15.5-26.0)	23.8 (18.4-28.6)	19.4 (14.1-24.9)†	23.0 (18.0-27.0)	24.0 (22.0-30.0)	21.0 (14.9-24.8)‡
Serum sodium, mmol/L	138 (136-141)	139 (137-141)	138 (135-140)‡	138 (136-140)	138 (137-141)	137 (135-139)†
Total cholesterol, mmol/L	4.1 (3.3-5.0)	4.3 (3.6-5.3)	3.8 (3.2-4.7)‡	4.0 (4.0-5.0)	4.5 (4.0-5.0)	4.0 (3.9-5.0)
Serum creatinine, µmol/L	103 (84-133)	95 (80-122)	110 (88-147)4‡	110 (80-136)	92 (74-118)	118 (94-156)‡
GFR, mL/min/1.73m^2	59 (43-74)	60 (47-77)	55 (40-70)‡	62 (46-85)	72 (55-92)	56 (40-68)‡
Urea, mmol/L	8.6 (6.7-12.0)	7.9 (6.3-10.6)	9.3 (7.2-13.3)‡	8.8 (6.6-11.9)	7.6 (5.6-9.4)	10.6 (7.7-13.8)‡
Uric acid, µmol/L	399 (320-493)	379 (321-457)	426 (318-515)†	435 (334-510)	392 (329-467)	462 (385-549)†
NT-proBNP, pg/mL	1832 (763-3945)	891 (290-1521)	3608 (1688-6088)‡	2608 (1377-5087)	2013 (1004-3551)	4025 (2164-6475)‡
LVEF, %	28 (23-33)	30 (25-34)	27 (22-32)‡	28 (24-33)	28 (23-32)	29 (24-33)
Medications						
ACE-I / ARB	1,197 (79.3)	551 (78.2)	646 (80.2)	146 (92.4)	76 (97.4)	70 (87.5)*
Beta-blocker	1,165 (77.1)	540 (76.6)	625 (77.6)	143 (90.5)	74 (94.9)	69 (86.3)
Furosemide	1,057 (70.0)	450 (63.8)	607 (75.4)‡	124 (78.5)	54 (69.2)	70 (87.5)†
Other loop diuretics	4 (0.3)	0 (0.0)	4 (0.5)	3 (1.9)	2 (2.6)	1 (1.3)
Thiazide diuretics	314 (20.8)	148 (21.0)	166 (20.6)	25 (15.8)	11 (14.1)	14 (17.5)
MRA	845 (56.0)	355 (50.4)	490 (60.9)‡	110 (69.6)	58 (74.4)	52 (65.0)
Digitalis	369 (24.4)	152 (21.6)	217 (27.0)*	46 (29.1)	19 (24.4)	27 (33.8)
Amiodarone	394 (26.1)	168 (23.8)	226 (28.1)	47 (29.7)	19 (24.4)	28 (35.0)
Statin	772 (51.1)	331 (47.0)	441 (54.8)†	97 (61.4)	49 (62.8)	48 (60.0)
Allopurinol	361 (23.9)	113 (16.0)	248 (30.8)‡	34 (21.5)	16 (20.5)	18 (22.5)

Continuous variables are presented as median (interquartile range), categorical variables as n (%).

**p < 0.05; †p < 0.01; ‡p < 0.001 vs. patients alive at 5-year follow-up within the same cohort, unpaired Student's t-test or Mann-Whitney U test for continuous variables, Chi-squared or Fisher's exact test for categorical variables.*

ACE-I – angiotensin-converting enzyme inhibitor, ARB – angiotensin II receptor blocker, CRT-D – cardiac resynchronization therapy-defibrillator, GFR – glomerular filtration rate, LBBB – left bundle branch block, LVEF – left ventricular ejection fraction, MRA – mineralocorticoid receptor antagonists, NT-proBNP – N-terminal pro-brain natriuretic peptide, NYHA – New York Heart Association functional class, SBP – systolic blood pressure

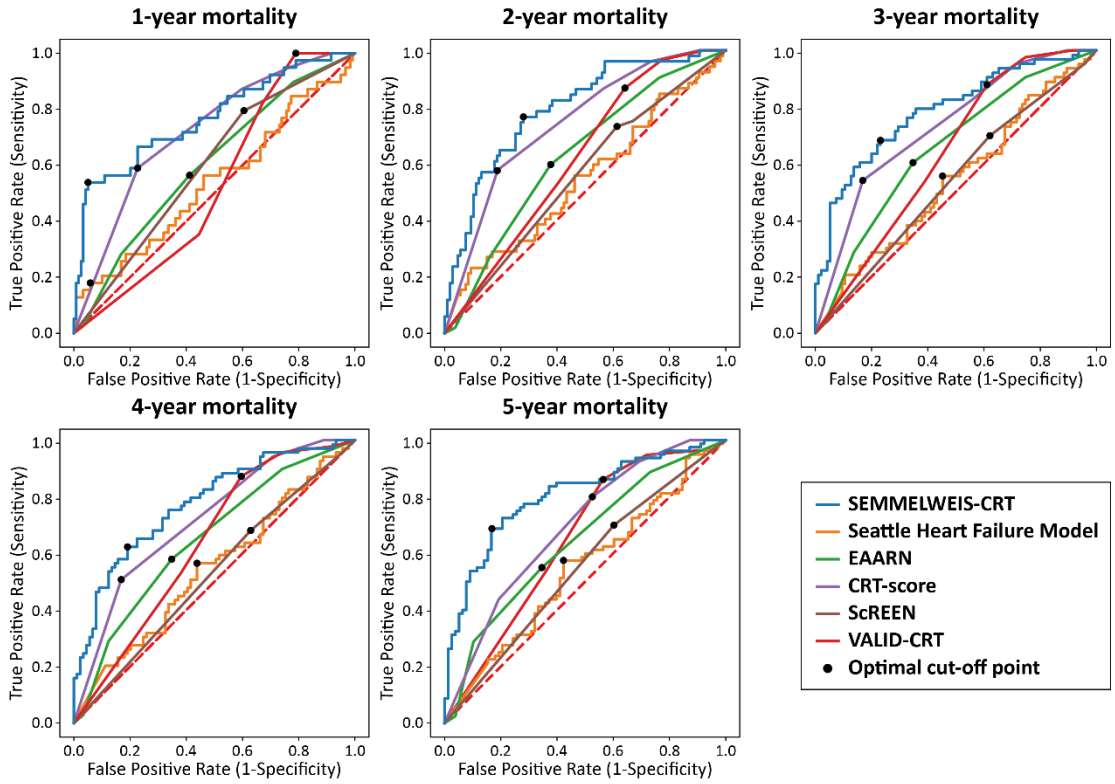


Figure 17 *Receiver operating characteristic curve analysis of the evaluated risk scores (166)*

Calibrated cumulative probabilities were used in the receiver operating characteristic curve analysis.

Table 8 *Area under the receiver operating characteristic curve of the different scores*

	1-year	2-year	3-year	4-year	5-year	Mean
SEMMELWEIS-CRT	0.768 (0.674-0.861)	0.793 (0.718-0.867)	0.785 (0.711-0.859)	0.776 (0.703-0.849)	0.803 (0.733-0.872)	0.785
SHFM	0.537 (0.426-0.647)*	0.543 (0.445-0.642)*	0.539 (0.447-0.632)*	0.544 (0.453-0.635)*	0.544 (0.454-0.634)*	0.541
EAARN	0.602 (0.505-0.699)*	0.627 (0.539-0.714)*	0.653 (0.570-0.736)*	0.649 (0.566-0.731)*	0.643 (0.560-0.726)*	0.635
VALID-CRT	0.529 (0.416-0.643)*	0.618 (0.523-0.713)*	0.638 (0.552-0.725)*	0.637 (0.550-0.724)*	0.650 (0.564-0.737)*	0.614
CRT-score	0.722 (0.637-0.806)	0.743 (0.667-0.818)	0.732 (0.657-0.807)	0.720 (0.644-0.795)	0.693 (0.615-0.771)*	0.722
ScREEN	0.595 (0.516-0.673)*	0.555 (0.477-0.633)*	0.536 (0.460-0.612)*	0.525 (0.449-0.601)*	0.549 (0.474-0.624)*	0.552

**p < 0.05 vs. SEMMELWEIS-CRT, DeLong test. Cell contents are areas under the receiver operating characteristic curves with 95% confidence intervals.*

SHFM – Seattle Heart Failure Model

Table 9 *Brier scores of the final model before and after calibration with Platt's scaling*

	1-year	2-year	3-year	4-year	5-year	Mean
Uncalibrated	0.178	0.186	0.215	0.267	0.335	0.236
Calibrated	0.164	0.180	0.201	0.208	0.207	0.192

Cell contents are Brier scores. Brier score is defined as the mean squared difference between the observed and predicted outcome. It ranges from 0 to 1, with values closer to 0 indicating better calibration.

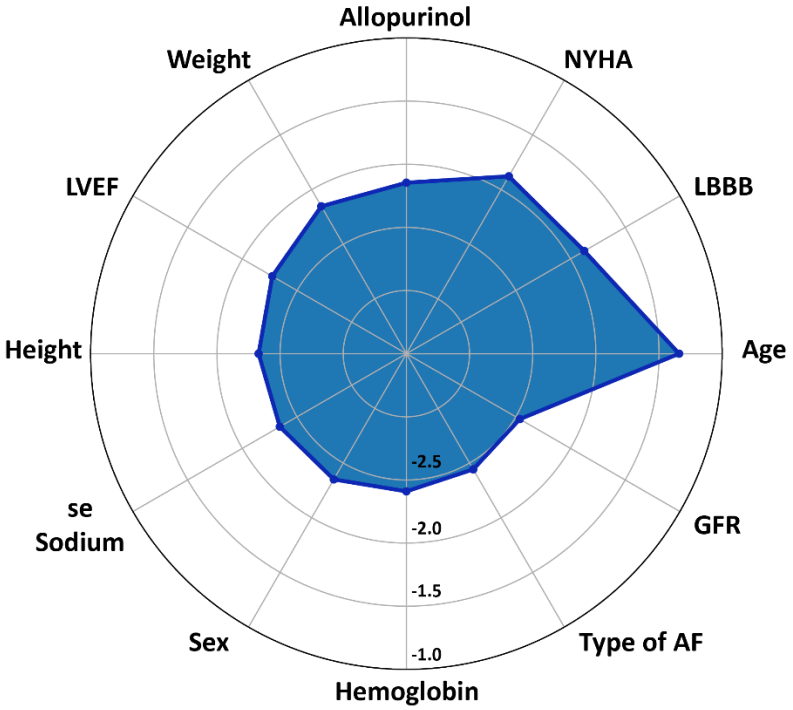


Figure 18 *The 12 most important predictors of all-cause mortality as assessed by the SEMMELWEIS-CRT score (166)*

The importance of each feature was quantified by calculating the decrease in the model's performance (area under the receiver operating characteristic curve) after permuting its values (permutation feature importances method). The higher its value, the more important the feature is. As the values of feature importances were spread over a wide range (more orders of magnitude), base-10 logarithmic transformation was performed to facilitate plotting.

AF – atrial fibrillation, GFR – glomerular filtration rate, LBBB – left bundle branch block, LVEF – left ventricular ejection fraction, NYHA – New York Heart Failure Association functional class

4.2.4 Machine learning-based risk stratification

Based on the predicted probability of death, patients were split into four quartiles at each year of follow-up. As depicted by Kaplan–Meier curves, there was a significant difference in the distribution of events across the quartiles at all years, and a graded increase in event rates could be observed while moving from the 2nd quartile through the 4th quartile (Figure 19). At 1-year follow-up, being categorized to the 4th quartile was associated with a more than 7-fold increased risk of death compared with those in the 1st quartile (Table 10). At 2-, 3-, 4-, and 5-year follow-up, patients in the 3rd and 4th quartiles exhibited a significantly increased risk of mortality compared with those in the 1st quartile (Table 10). The expected survival of patients was monotonously decreasing from the 1st through the 4th quartile in each year (Table 11).

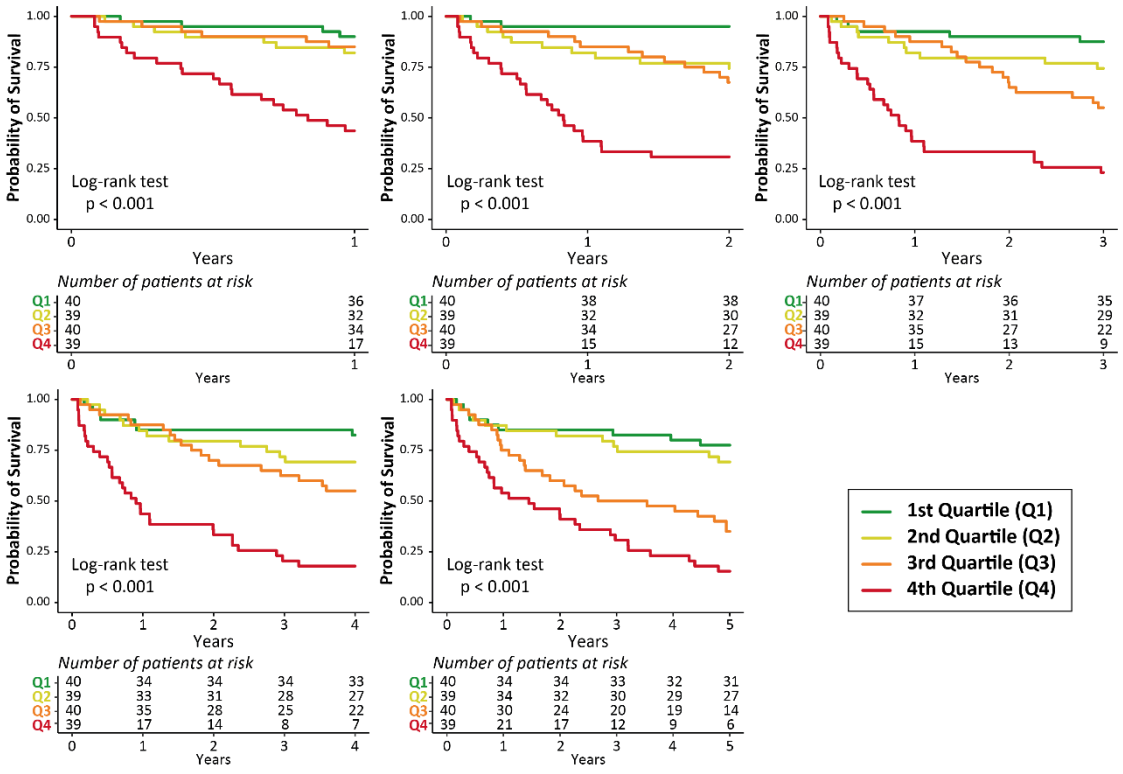


Figure 19 *Survival analysis of the quartiles (166)*

Based on the predicted probability of death, patients were split into four quartiles at each year of follow-up. The survival of the quartiles was visualized on Kaplan–Meier curves, and log-rank tests were performed for comparison.

Table 10 *Hazard ratios of all-cause mortality in different quartiles*

	1-year	2-year	3-year	4-year	5-year
2nd vs. 1st quartile	1.89 (0.55-6.45) p=0.301	5.55 (1.22-25.35) p=0.010	2.18 (0.44-3.36) p=0.142	1.81 (0.717-4.60) p=0.203	1.40 (0.59-3.33) p=0.439
3rd vs. 1st quartile	1.56 (0.44-5.52) p=0.487	7.30 (1.65-32.37) p<0.001	4.18 (1.55-11.28) p=0.002	2.88 (1.20-6.90) p=0.012	3.75 (1.75-8.04) p<0.001
4th vs. 1st quartile	7.92 (2.72-23.07) p<0.001	21.55 (5.10-91.06) p<0.001	10.59 (4.07-27.56) p<0.001	8.16 (3.56-18.72) p<0.001	6.71 (3.17-14.21) p<0.001

Patients were split (repeatedly) into four quartiles based on the predicted probability of death in each year. As the quartiles in each year might contain a different set of patients, row-wise evaluation of hazard ratios should be avoided. Cell contents are hazard ratios (95% confidence interval) with p-values calculated using Cox proportional hazards models.

Table 11 *The average expected survival time in each quartile*

	1-year	2-year	3-year	4-year	5-year
1st quartile	3.47 ± 0.42	3.50 ± 0.41	3.50 ± 0.43	3.53 ± 0.38	3.50 ± 0.41
2nd quartile	2.86 ± 0.37	2.86 ± 0.27	2.88 ± 0.23	2.85 ± 0.22	2.83 ± 0.29
3rd quartile	2.39 ± 0.29	2.39 ± 0.26	2.36 ± 0.23	2.33 ± 0.26	2.30 ± 0.33
4th quartile	1.78 ± 0.32	1.74 ± 0.28	1.76 ± 0.30	1.79 ± 0.34	1.86 ± 0.43

Cell contents are mean ± standard deviation of the expected survival (years).

4.3 Exploring the sex-specific differences and similarities in the predictors of mortality among patients undergoing CRT implantation

4.3.1 Baseline clinical characteristics and all-cause mortality

The final 1- and 3-year cohorts included 2,191 (74.7% males, 56.7% CRT-D) and 1,900 patients (75.0% males, 54.1% CRT-D), respectively (Figure 20). In the 1-year cohort, 50.4% of the patients had ischemic etiology of HF, 57.8% had NYHA functional class III/IV, and the median LVEF was 28 (24 – 32) %. In the 3-year cohort, ischemic etiology was reported in 51.5% of the patients, 61.0% presented with NYHA functional class III/IV, and the median LVEF was 28 (24 – 32) %. The baseline clinical characteristics of the patients are summarized in Tables 12 and 13.

In the 1-year cohort, 203 (12.4%) men and 49 (8.8%) women died during the 1-year follow-up period. Univariable Cox regression analysis revealed a significantly lower risk of all-cause mortality in women compared to men (HR: 0.698, 95% CI: 0.511 – 0.954; $p = 0.024$); however, after adjusting for age, etiology of HF, QRS morphology, type of the implanted device, and type of AF (history of or current), we could not observe a significant difference between sexes (HR: 0.803, 95% CI: 0.581 – 1.110; $p = 0.183$) (Figure 21A).

As observed in the 1-year cohort, males exhibited significantly higher mortality rates compared to females in the 3-year cohort as well (502 [35.2%] vs. 113 [23.8%]; $p < 0.001$). The univariable Cox regression analysis also confirmed this finding as it showed a significantly lower risk of all-cause mortality in females compared to males (HR: 0.625, 95% CI: 0.510 – 0.767; $p < 0.001$) (Figure 21B). Moreover, this difference remained significant even after adjusting for the previously listed covariates (HR: 0.686, 95% CI: 0.555 – 0.848; $p < 0.001$).

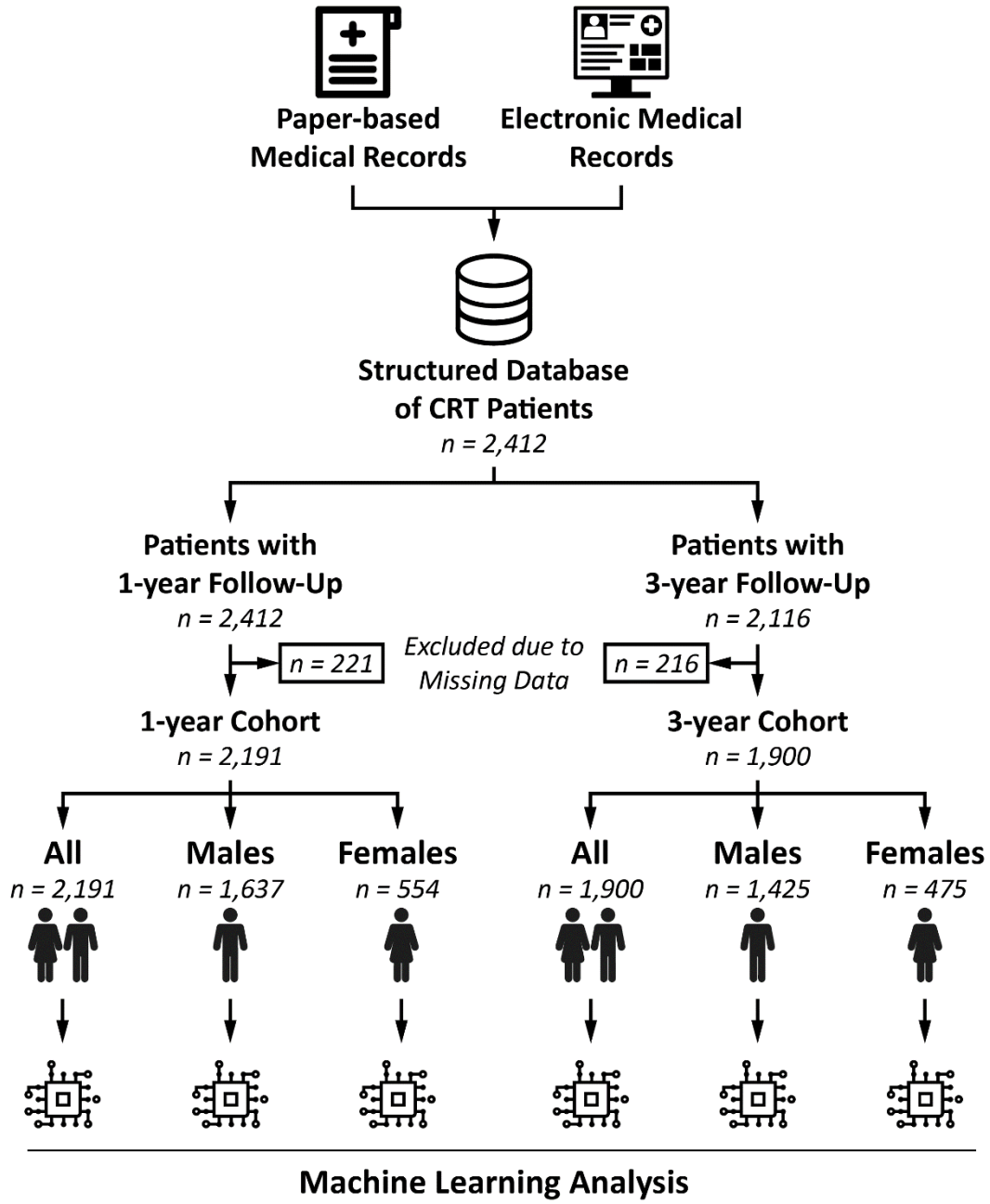


Figure 20 *Flowchart illustrating the steps of patient selection (173)*

For each patient who underwent successful CRT implantation at our center, pre-implantation clinical characteristics and procedural parameters were collected retrospectively from paper-based or electronic medical records and entered into our structured database. After excluding patients with ≥30% missing values, machine learning models were implemented to predict 1- and 3-year all-cause mortality in the entire cohort, in males and females separately (altogether 6 separate binary classification tasks).

CRT – cardiac resynchronization therapy

Table 12 *Clinical characteristics of the 1-year cohort*

	All patients (n=2,191)	Males (n=1,637)	Females (n=554)	p-value
Demographics, vitals, and key electrophysiological characteristics				
Age, years*	68 (61-74)	68 (60-74)	69 (63-75)	<0.001
Weight, kg (1423)	80 (70-91)	84 (75-95)	70 (60-80)	<0.001
Height, cm (1413)	172 (165-177)	175 (170-179)	162 (157-167)	<0.001
BMI, kg/m^2 (1413)*	27.4 (24.5-30.7)	27.6 (24.8-30.8)	26.7 (23.4-30.5)	<0.001
SBP, mmHg (807)	125 (111-136)	125 (111-136)	124 (110-136)	0.403
DBP, mmHg (807)	73 (65-80)	74 (65-80)	71 (64-80)	0.089
NYHA III/IV (1803)*	1,043 (57.8)	781 (57.9)	262 (57.7)	0.945
CRT-D*	1,239 (56.5)	1,005 (61.4)	234 (42.2)	<0.001
QRS duration, ms (754)	160 (140-180)	160 (140-180)	160 (140-170)	0.068
QRS morphology, LBBB*	1,572 (71.7)	1,127 (68.8)	445 (80.3)	<0.001
LV lead position (1890)*				
Anterior	84 (4.4)	62 (4.4)	22 (4.7)	
Lateral	1,227 (64.9)	932 (65.7)	295 (62.5)	
Posterior	579 (30.6)	424 (25.9)	155 (32.8)	0.442
Medical history				
Ischemic etiology of HF*	1,104 (50.4)	902 (55.1)	202 (36.5)	<0.001
History of MI	868 (39.6)	713 (43.6)	155 (28.0)	<0.001
HF duration >18 months*	680 (31.0)	519 (31.7)	161 (29.1)	0.245
History of or current AF*				
No AF	1,394 (63.6)	998 (61.0)	396 (71.5)	
Paroxysmal	342 (15.6)	257 (15.7)	85 (15.3)	
Persistent	59 (2.7)	51 (3.1)	8 (1.4)	
Permanent	396 (18.1)	331 (20.2)	65 (11.7)	<0.001
Valvular heart disease*	135 (6.2)	99 (6.0)	36 (6.5)	0.780
Hypertension*	1,618 (73.8)	1,216 (74.3)	402 (72.6)	0.459
Diabetes mellitus*	813 (37.1)	624 (38.1)	189 (34.1)	0.092
COPD*	325 (14.8)	239 (14.6)	86 (15.5)	0.597
Current smoker*	131 (6.0)	103 (6.3)	28 (5.1)	0.288
Laboratory parameters				
Hemoglobin, g/L (1440)*	136 (123-148)	139 (126-150)	130 (120-140)	<0.001
Serum sodium, mmol/L (1374)*	138 (136-141)	138 (136-140)	139 (136-141)	0.019
Total cholesterol, mmol/L (956)	4.1 (3.4-5.1)	4.0 (3.3-4.9)	4.7 (3.6-5.5)	<0.001
Serum creatinine, µmol/L (1473)*	101 (82-131)	105 (87-134)	86 (71-112)	<0.001
Urea, mmol/L (1445)	8.3 (6.4-11.7)	8.6 (6.6-11.8)	7.5 (6.0-10.9)	<0.001
Uric acid, µmol/L (766)	405 (322-492)	412 (330-494)	383 (307-474)	0.020
NT-proBNP, pg/mL (309)	2640 (1262–3699)	2490 (1367–3473)	2680 (1250–3710)	0.938
Echocardiographic parameters				
LV ejection fraction, % (1610)*	28 (24-32)	28 (23-32)	28 (25-33)	0.046
LVEDD, mL (1610)*	64 (58-70)	65 (59-71)	61 (55-66)	<0.001
Medications				
ACE-I / ARB*	2,014 (91.9)	1,509 (92.2)	505 (91.2)	0.499
Beta-blocker*	1,951 (89.0)	1,457 (89.0)	494 (89.2)	0.914
Ca-channel blocker*	127 (5.8)	99 (6.0)	28 (5.1)	0.387
Loop diuretics*	1,757 (80.2)	1,315 (80.3)	442 (79.8)	0.780
Thiazide diuretics*	516 (23.6)	402 (24.6)	114 (20.6)	0.056
MRA*	1,497 (68.3)	1,115 (68.1)	382 (69.0)	0.713
Digitalis*	464 (21.2)	359 (21.9)	105 (19.0)	0.138
Amiodarone*	593 (27.1)	466 (28.5)	127 (22.9)	0.011
Statin*	1,314 (60.0)	995 (60.8)	319 (57.6)	0.184
Allopurinol*	591 (27.0)	475 (29.0)	116 (20.9)	<0.001
Oral anticoagulants	729 (33.3)	598 (36.5)	131 (23.6)	<0.001
1-year all-cause mortality	252 (11.5)	203 (12.4)	49 (8.8)	0.028

**Features included in the machine learning models.*

The value (in parenthesis) after a feature's name indicates the number of patients with available data. If there is no value reported, the given feature was available for all patients.

Table 12 *(continued)*

Continuous variables are presented as median (interquartile range), categorical variables as n (%). The comparison between males and females was performed using unpaired Student's t-test or Mann-Whitney U test for continuous variables, Chi-squared or Fisher's exact test for categorical variables, as appropriate.

ACE-I – angiotensin-converting enzyme inhibitors, AF – atrial fibrillation, ARB – angiotensin receptor blockers, BMI – body mass index, COPD – chronic obstructive pulmonary disease, CRT-D – cardiac resynchronization therapy defibrillator; DBP – diastolic blood pressure, DM, diabetes mellitus; HF – heart failure, LBBB – left bundle branch block, LVEDD – left ventricular end-diastolic diameter, MI – myocardial infarction, MRA – mineralocorticoid receptor antagonists, NT-proBNP – N-terminal pro-brain natriuretic peptide, NYHA – New York Heart Association functional class, SBP – systolic blood pressure

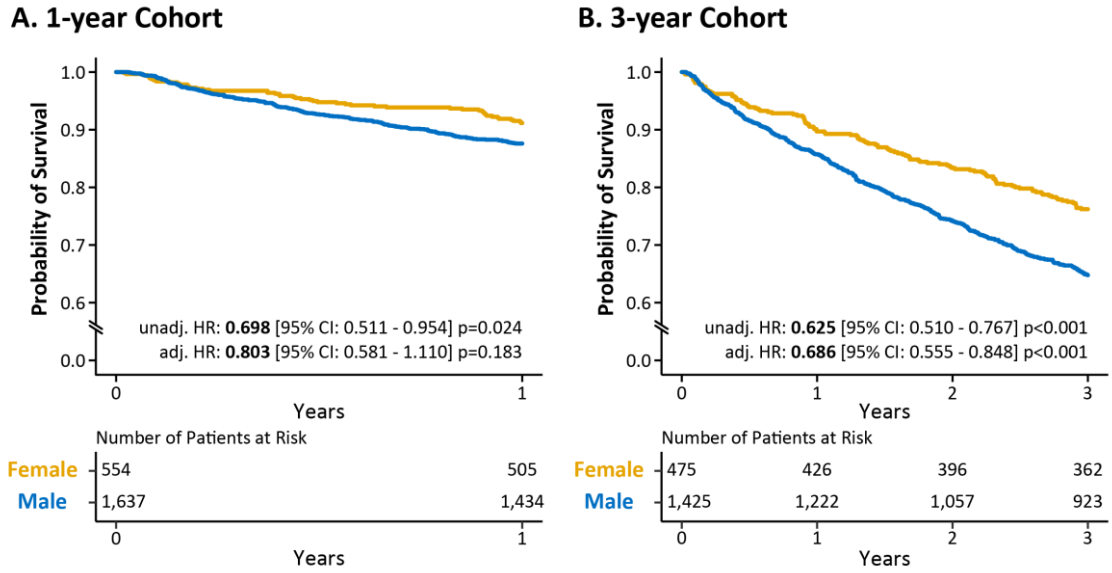


Figure 21 *Kaplan-Meier curves for males and females in the 1- (A) and 3-year (B) cohorts (173)*

Kaplan-Meier curve analysis illustrates the difference in the survival of male and female CRT patients during 1- and 3-year follow-up. Cox proportional hazards models were used to compute hazard ratios with 95% confidence intervals. Hazard ratios were adjusted for age (at implantation), QRS morphology, etiology of heart failure, the type of the implanted device, and the type of atrial fibrillation.

CI – confidence interval, CRT – cardiac resynchronization therapy, HR – hazard ratio

Table 13 *Clinical characteristics of the 3-year cohort*

	All patients (n=1,900)	**Males (n=1,425)**	**Females (n=475)**	**p-value**
Demographics, vitals, and key electrophysiological characteristics				
Age, years*	68 (61-74)	68 (60-74)	69 (63-75)	<0.001
Weight, kg (1280)	80 (70-90)	84 (75-95)	70 (60-80)	<0.001
Height, cm (1270)	172 (165-177)	175 (170-179)	161 (157-167)	<0.001
BMI, kg/m^2 (1270)*	27.3 (24.3-30.5)	27.5 (24.7-30.5)	26.5 (23.3-30.5)	<0.001
SBP, mmHg (660)	123 (110-136)	124 (111-136)	122 (110-135)	0.463
DBP, mmHg (660)	72 (65-80)	72 (65-80)	71 (64-80)	0.292
NYHA III/IV (1568)*	956 (61.0)	719 (61.0)	237 (60.9)	0.984
CRT-D*	1027 (54.1)	839 (58.9)	188 (39.6)	<0.001
QRS duration, ms (718)	160 (140-180)	160 (142-180)	160 (140-170)	0.035
QRS morphology, LBBB*	1,385 (72.9)	1,000 (70.2)	385 (81.1)	<0.001
LV lead position (1630)*				
Anterior	75 (4.6)	54 (4.4)	21 (5.2)	
Lateral	1072 (65.8)	814 (66.3)	258 (64.0)	
Posterior	483 (29.6)	359 (29.3)	124 (30.8)	0.633
Medical history				
Ischemic etiology*	979 (51.5)	802 (56.3)	177 (37.3)	<0.001
History of MI	793 (41.7)	655 (46.0)	138 (29.1)	<0.001
HF duration >18 months*	616 (32.4)	477 (33.5)	139 (29.3)	0.090
History of or current AF*				
No AF	1,181 (62.2)	850 (59.6)	331 (69.7)	
Paroxysmal	306 (16.1)	227 (15.9)	79 (16.6)	
Persistent	49 (2.6)	43 (3.0)	6 (1.3)	
Permanent	364 (19.2)	305 (21.4)	59 (12.4)	<0.001
Valvular heart disease*	131 (6.9)	97 (6.8)	34 (7.2)	0.875
Hypertension*	1,417 (74.6)	1,067 (74.9)	350 (73.7)	0.648
Diabetes mellitus*	704 (37.1)	542 (38.0)	162 (34.1)	0.125
COPD*	288 (15.2)	213 (14.9)	75 (15.8)	0.658
Current smoker*	110 (5.8)	89 (6.2)	21 (4.4)	0.140
Laboratory parameters				
Hemoglobin, g/L (1254)*	136 (123-148)	139 (125-150)	131 (120-140)	<0.001
Serum sodium, mmol/L (1180)*	138 (136-141)	138 (136-140)	139 (136-141)	0.020
Total cholesterol, mmol/L (827)	4.1 (3.4-5.1)	4 (3.3-4.9)	4.7 (3.6-5.5)	<0.001
Serum creatinine, µmol/L (1278)*	102 (82-132)	106 (87-135)	87 (71-113)	<0.001
Urea, mmol/L (1254)	8.5 (6.4-11.7)	8.8 (6.6-12.0)	7.7 (6.1-10.9)	<0.001
Uric acid, µmol/L (655)	406 (323-494)	409 (329-495)	386 (313-479)	0.082
NT-proBNP, pg/mL (237)	2758 (1398–3570)	2610 (1496–3376)	2804 (1290–3616)	0.931
Echocardiographic parameters				
LV ejection fraction, % (1378)*	28 (24-32)	28 (23-32)	28 (25-32)	0.185
LVEDD, mL (1378)*	64 (58-70)	65 (59-71)	61 (56-67)	<0.001
Medications				
ACE-I / ARB*	1,731 (91.1)	1,303 (91.4)	428 (90.1)	0.429
Beta-blocker*	1,691 (89.0)	1,264 (88.7)	427 (89.9)	0.472
Ca-channel blocker*	106 (5.6)	81 (5.7)	25 (5.3)	0.729
Loop diuretics*	1,526 (80.3)	1,153 (80.9)	373 (78.5)	0.257
Thiazide diuretics*	456 (24.0)	354 (24.8)	102 (21.5)	0.137
MRA*	1,270 (66.8)	953 (66.9)	317 (66.7)	0.955
Digitalis*	442 (23.3)	341 (23.9)	101 (21.3)	0.234
Amiodarone*	528 (27.8)	415 (29.1)	113 (23.8)	0.025
Statin*	1,134 (59.7)	862 (60.5)	272 (57.3)	0.214
Allopurinol*	521 (27.4)	422 (29.6)	99 (20.8)	<0.001
Oral anticoagulants	627 (33.0)	510 (35.8)	117 (24.6)	<0.001
3-year all-cause mortality	615 (32.4)	502 (35.2)	113 (23.8)	<0.001

**Features included in the machine learning models.*

The value (in parenthesis) after a feature's name indicates the number of patients with available data. If there is no value reported, the given feature was available for all patients.

Table 13 *(continued)*

Continuous variables are presented as median (interquartile range), categorical variables as n (%). The comparison between males and females was performed using unpaired Student's t-test or Mann-Whitney U test for continuous variables, Chi-squared or Fisher's exact test for categorical variables, as appropriate.

ACE-I – angiotensin-converting enzyme inhibitors, AF – atrial fibrillation, ARB – angiotensin receptor blockers, BMI – body mass index, COPD – chronic obstructive pulmonary disease, CRT-D – cardiac resynchronization therapy defibrillator; DBP – diastolic blood pressure, DM, diabetes mellitus; HF – heart failure, LBBB – left bundle branch block, LVEDD – left ventricular end-diastolic diameter, MI – myocardial infarction, MRA – mineralocorticoid receptor antagonists, NT-proBNP – N-terminal pro-brain natriuretic peptide, NYHA – New York Heart Association functional class, SBP – systolic blood pressure

4.3.2 Machine learning for the prediction of 1- and 3-year all-cause mortality

Among the evaluated ML classifiers, CIRF exhibited the best performance for discrimination between survival/all-cause death with an AUC of 0.717 (95% CI: 0.676 – 0.758) and 0.739 (95% CI: 0.715 – 0.762) in the 1- and 3-year training cohorts, respectively. When evaluating the models' discriminatory power in the test cohorts, we observed an AUC of 0.728 (95% CI: 0.645 – 0.802) and 0.732 (95% CI: 0.681 – 0.784) for the prediction of 1- and 3-year mortality, respectively. Models were also trained and tested separately in the female and male subsets of the 1- and 3-year cohorts. The AUCs ranged from 0.712 to 0.748 in the training sets and from 0.681 to 0.798 in the test sets suggesting a modest variability in the models' predictive capabilities across the different subsets of patients (Tables 14 and 15).

After sorting the patients in ascending order based on the predicted probability of death and plotting the distribution of probability values, the accumulation of patients who died during the given follow-up period could be observed in the higher risk regions of the plots (Figure 22). These findings suggest that our models can perform risk stratification effectively.

The Brier score – measuring the accuracy of the probabilistic predictions – for the 1- and 3-year models were 0.197 and 0.201, indicating a sufficiently good calibration of our models. Table 16 summarizes the Brier scores for the remainder of the CIRF models.

Table 14 *Training and testing of the conditional inference random forest models in the entire 1-year cohort (n=2,191), males (n=1,637), and females (n=554)*

	Hyperparameter space	Best combination of hyperparameters	AUC in the training cohort (10-fold CV)	AUC in the test cohort
Entire 1-year cohort	('alpha': [0.05, 0.1], 'max_depth': [-1, 2, 16], 'min_samples_split': [2, 4, 10], 'n_estimators': [10, 25, 50, 100, 200, 400], 'n_permutations': [50, 100, 150])	('alpha': 0.1, 'max_depth': -1, 'min_samples_split': 2, 'n_estimators': 100, 'n_permutations': 50)	0.717 (0.676-0.758)	0.728 (0.654-0.802)
Female		('alpha': 0.1, 'max_depth': -1, 'min_samples_split': 2, 'n_estimators': 50, 'n_permutations': 50)	0.748 (0.684-0.812)	0.798 (0.691-0.905)
Male		('alpha': 0.1, 'max_depth': -1, 'min_samples_split': 10, 'n_estimators': 200, 'n_permutations': 50)	0.712 (0.690-0.735)	0.697 (0.610-0.783)

Hyperparameters not listed in the table were set to default values predefined in the scikit-learn module (version 0.21.3). Hyperparameter tuning was performed with grid search. Models with each combination of hyperparameters were trained with stratified 10-fold cross-validation in the training cohorts, and the average AUC was calculated. Using the best performing combination of hyperparameters (3rd and 4th columns), models were evaluated in the test cohorts of each subset as well (5th column).

AUC – area under the receiver operating characteristic curve, CV – cross-validation

Table 15 *Training and testing of the conditional inference random forest models in the entire 3-year cohort (n=1,900), males (n=1,425), and females (n=475)*

	Hyperparameter space	Best combination of hyperparameters	AUC in the training cohort (10-fold CV)	AUC in the test cohort
Entire 3-year cohort	('alpha': [0.05, 0.1], 'max_depth': [-1, 2, 16], 'min_samples_split': [2, 4, 10], 'n_estimators': [10, 25, 50, 100, 200, 400], 'n_permutations': [50, 100, 150])	('alpha': 0.1, 'max_depth': -1, 'min_samples_split': 4, 'n_estimators': 400, 'n_permutations': 100)	0.739 (0.715-0.762)	0.732 (0.681-0.784)
Female		('alpha': 0.1, 'max_depth': -1, 'min_samples_split': 10, 'n_estimators': 400, 'n_permutations': 150)	0.719 (0.678-0.760)	0.777 (0.678-0.876)
Male		('alpha': 0.1, 'max_depth': -1, 'min_samples_split': 4, 'n_estimators': 200, 'n_permutations': 150)	0.740 (0.717-0.763)	0.681 (0.617-0.745)

Hyperparameters not listed in the table were set to default values predefined in the scikit-learn module (version 0.21.3). Hyperparameter tuning was performed with grid search. Models with each combination of hyperparameters were trained with stratified 10-fold cross-validation in the training cohorts, and the average AUC was calculated. Using the best performing combination of hyperparameters (3rd and 4th columns), models were evaluated in the test cohorts of each subset as well (5th column).

AUC – area under the receiver operating characteristic curve, CV – cross-validation

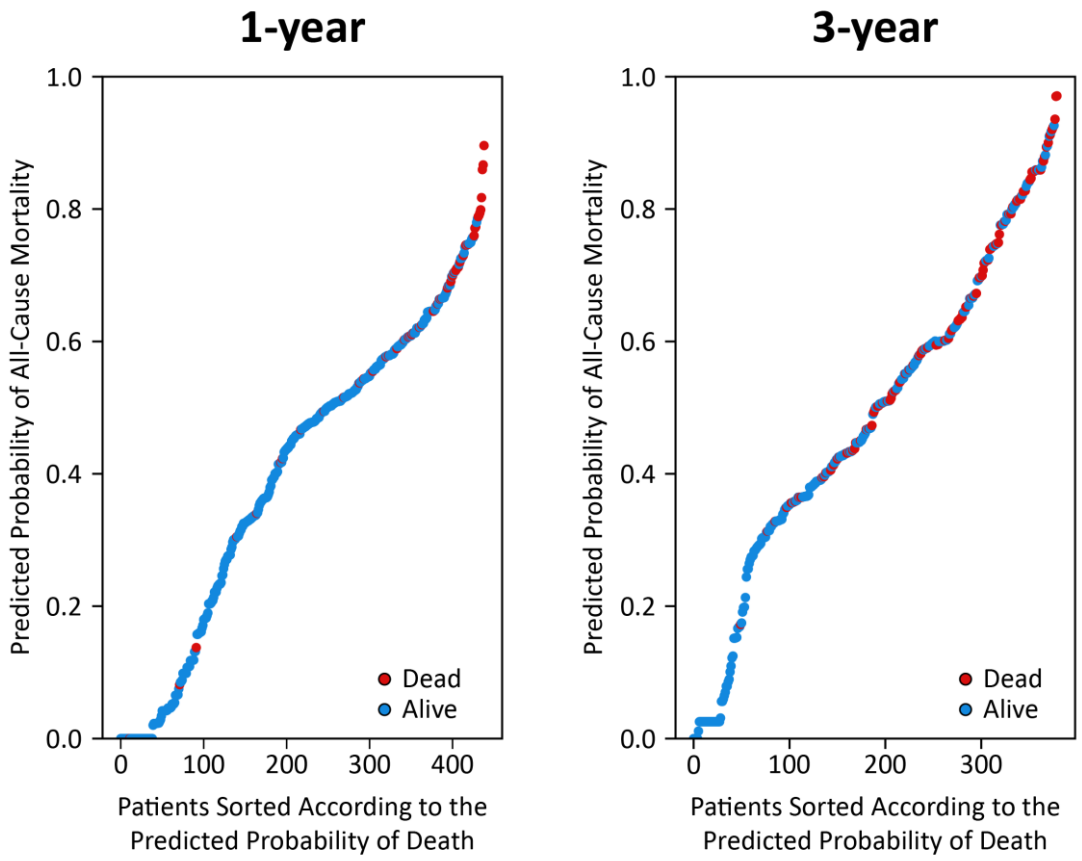


Figure 22 *Distribution of the predicted probabilities in the 1- and 3-year test cohorts (173)*

Patients were sorted in ascending order based on the predicted probability of death, and the distribution of probability values was plotted. The accumulation of patients who died during the given follow-up period could be observed in the higher risk regions of the plots in the 1- and 3-year cohorts as well.

CRT – cardiac resynchronization therapy

Table 16 *Brier scores of the final machine learning models*

	1-year test cohorts	3-year test cohorts
All	0.197	0.201
Females	0.193	0.203
Males	0.199	0.211

Cell contents are Brier scores (calculated using the probabilities of the test cohorts). Brier score is defined as the mean squared difference between the observed and predicted outcome. It ranges from 0 to 1, with values closer to 0 indicating better calibration.

4.3.3 Most important predictors of mortality as assessed using machine learning

Leading predictors of all-cause mortality are illustrated in Figure 23, and the comprehensive list of feature importances is provided in Tables 17 and 18.

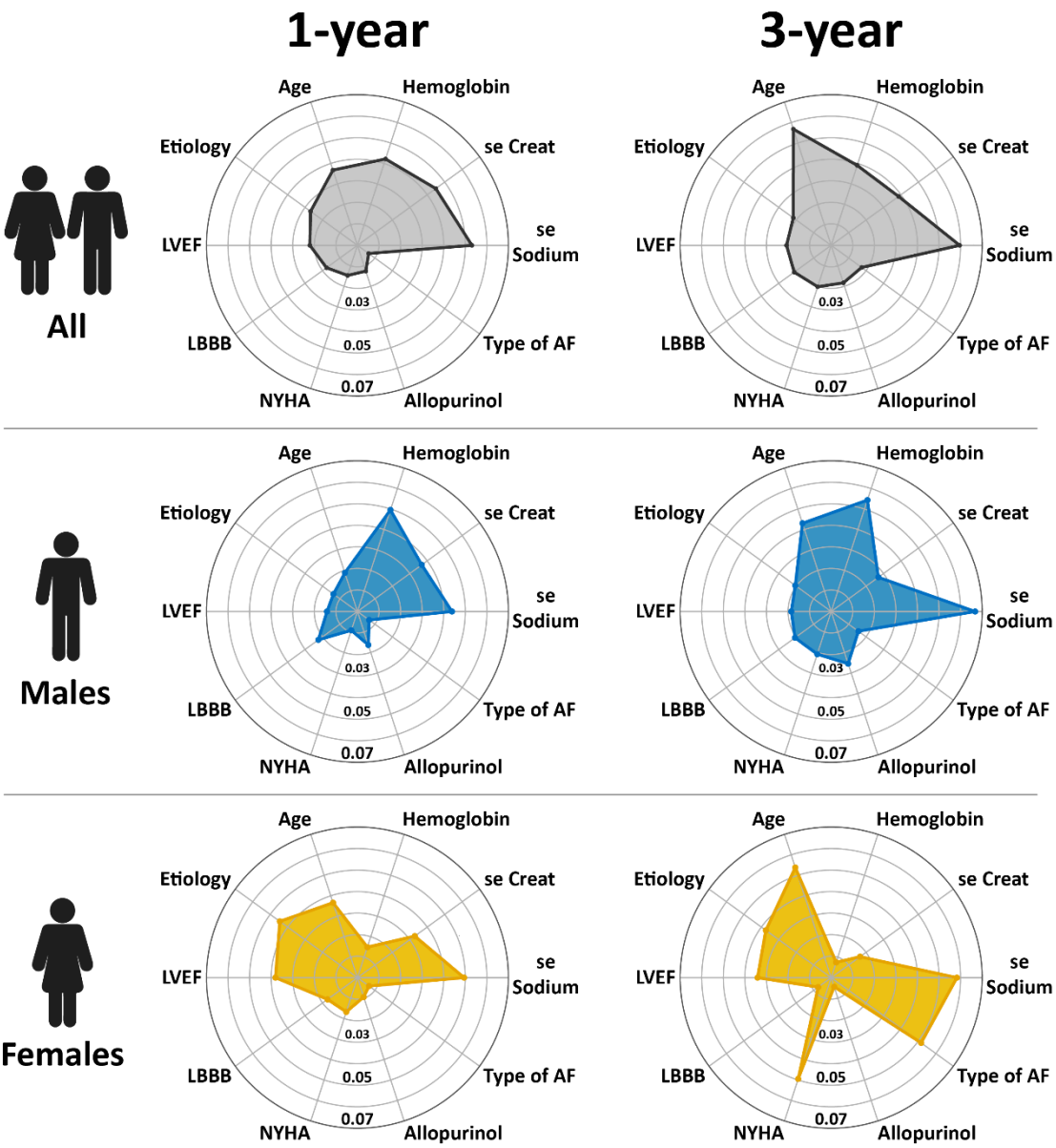


Figure 23 *The most important predictors of 1- and 3-year all-cause mortality in patients undergoing CRT implantation (173)*

The importance of each feature was quantified with the permutation feature importances method, which measures the importance of a feature by calculating the mean decrease in the model's performance (area under the receiver operating characteristic curve) after permuting its values 10 times (see text for further details). To keep the data comparable between the different models, we identified the top 5 predictors in each model and took the union of these features; then, we plotted the results on radar charts.

AF – atrial fibrillation, LBBB – left bundle branch block, LVEF – left ventricular ejection fraction, NYHA – New York Heart Association

Table 17 *Permutation feature importances of the input variables calculated in the 1-year training cohorts*

All		Males		Females	
Feature	Importance	Feature	Importance	Feature	Importance
Serum Sodium	0.0531	Hemoglobin	0.0496	Serum Sodium	0.0495
Serum Creatinine	0.0449	Serum Sodium	0.0438	Etiology of HF	0.0445
Hemoglobin	0.0422	Serum Creatinine	0.0369	LV ejection fraction	0.0380
Age at CRT implantation	0.0368	QRS morphology - LBBB	0.0224	Age at CRT implantation	0.0366
Etiology of HF	0.0269	Age at CRT implantation	0.0189	Serum Creatinine	0.0329
Body Mass Index	0.0236	Allopurinol	0.0163	LV end-diastolic diameter	0.0216
LV ejection fraction	0.0221	Type of device (CRT-P or CRT-D)	0.0146	Body Mass Index	0.0197
QRS morphology - LBBB	0.0177	Body Mass Index	0.0141	QRS morphology - LBBB	0.0172
LV end-diastolic diameter	0.0175	LV ejection fraction	0.0141	NYHA functional class	0.0167
Loop diuretics	0.0173	Etiology of HF	0.0138	Amiodarone	0.0157
Type of device (CRT-P or CRT-D)	0.0166	LV end-diastolic diameter	0.0114	Hemoglobin	0.0149
NYHA functional class	0.0147	Diabetes	0.0096	Statin	0.0132
Statin	0.0132	Beta blockers	0.0094	MRA	0.0100
Allopurinol	0.0126	Loop diuretics	0.0094	Diabetes	0.0100
MRA	0.0124	NYHA functional class	0.0092	Thiazide diuretics	0.0098
COPD	0.0119	MRA	0.0084	Allopurinol	0.0095
HF duration >18 months	0.0109	Valvular heart disease	0.0082	Loop diuretics	0.0094
LV lead position	0.0103	Thiazide diuretics	0.0082	LV lead position	0.0083
Gender	0.0098	Amiodarone	0.0077	Valvular heart disease	0.0079
Amiodarone	0.0096	COPD	0.0071	Hypertension	0.0075
Digitalis	0.0088	Type of AF	0.0067	HF duration >18 months	0.0074
Diabetes	0.0088	HF duration >18 months	0.0063	Digitalis	0.0074
Thiazide diuretics	0.0081	Statin	0.0061	Type of AF	0.0066
Type of AF	0.0065	LV lead position	0.0055	COPD	0.0060
Beta blockers	0.0064	Hypertension	0.0053	ACE-I / ARB	0.0059
Hypertension	0.0057	Digitalis	0.0048	Type of device (CRT-P or CRT-D)	0.0047
Valvular heart disease	0.0053	Smoking status	0.0031	Beta blockers	0.0038
ACE-I / ARB	0.0051	ACE-I / ARB	0.0021	Smoking status	0.0003
Smoking status	0.0019	Calcium channel blockers	0.0008	Calcium channel blockers	0.0003
Calcium channel blockers	0.0013				

The importance of each feature was quantified with the permutation feature importances method, which measures the importance of a feature by calculating the mean decrease in the model's performance (area under the ROC curve) after permuting its values 10 times. Features are sorted according to permutation importance.

ACE-I – angiotensin-converting enzyme inhibitors, AF – atrial fibrillation, ARB - angiotensin II receptor blockers, COPD– chronic obstructive pulmonary disease, CRT – cardiac resynchronization therapy, HF – heart failure, LBBB – left bundle branch block, LV – left ventricular, MRA – mineralocorticoid receptor antagonists, NYHA – New York Heart Association functional class, ROC – receiver operating characteristic

Table 18 *Permutation feature importances of the input variables calculated in the 3-year training cohorts*

All		Males		Females	
Feature	Importance	Feature	Importance	Feature	Importance
Serum Sodium	0.0596	Serum Sodium	0.0667	Serum Sodium	0.0583
Age at CRT implantation	0.0568	Hemoglobin	0.0544	Age at CRT implantation	0.0538
Hemoglobin	0.0391	Age at CRT implantation	0.0431	Type of AF	0.0516
Serum Creatinine	0.0387	Serum Creatinine	0.0270	NYHA functional class	0.0495
Etiology of HF	0.0217	Allopurinol	0.0255	Etiology of HF	0.0374
QRS morphology - LBBB	0.0212	NYHA functional class	0.0208	LV ejection fraction	0.0341
Body Mass Index	0.0208	QRS morphology - LBBB	0.0207	MRA	0.0238
LV ejection fraction	0.0207	Etiology of HF	0.0206	Hypertension	0.0216
NYHA functional class	0.0202	LV ejection fraction	0.0185	Body Mass Index	0.0193
Digitalis	0.0187	Diabetes	0.0169	Serum Creatinine	0.0166
Allopurinol	0.0183	LV end-diastolic diameter	0.0166	Thiazide diuretics	0.0144
LV end-diastolic diameter	0.0182	Body Mass Index	0.0166	Digitalis	0.0127
Gender	0.0175	MRA	0.0160	LV end-diastolic diameter	0.0093
Type of AF	0.0175	Type of AF	0.0156	Type of device (CRT-P or CRT-D)	0.0084
Loop diuretics	0.0150	Loop diuretics	0.0156	COPD	0.0075
Type of device (CRT-P or CRT-D)	0.0118	Digitalis	0.0154	Hemoglobin	0.0075
MRA	0.0111	Type of device (CRT-P or CRT-D)	0.0149	QRS morphology - LBBB	0.0075
HF duration >18 months	0.0106	Hypertension	0.0134	Statin	0.0071
Diabetes	0.0102	LV lead position	0.0113	Allopurinol	0.0045
Statin	0.0096	Statin	0.0102	HF duration >18 months	0.0043
LV lead position	0.0095	HF duration >18 months	0.0098	LV lead position	0.0040
Thiazide diuretics	0.0091	Thiazide diuretics	0.0095	Diabetes	0.0036
COPD	0.0082	Amiodarone	0.0074	Amiodarone	0.0035
Hypertension	0.0066	COPD	0.0062	ACE-I / ARB	0.0029
Amiodarone	0.0059	ACE-I / ARB	0.0035	Loop diuretics	0.0017
ACE-I / ARB	0.0039	Beta blockers	0.0033	Valvular heart disease	0.0016
Beta blockers	0.0034	Smoking status	0.0029	Beta blockers	0.0001
Valvular heart disease	0.0025	Valvular heart disease	0.0027	Smoking status	0.0001
Calcium channel blockers	0.0017	Calcium channel blockers	0.0016	Calcium channel blockers	0.0001
Smoking status	0.0010				

The importance of each feature was quantified with the permutation feature importances method, which measures the importance of a feature by calculating the mean decrease in the model's performance (area under the ROC curve) after permuting its values 10 times. Features are sorted according to permutation importance.

ACE-I – angiotensin-converting enzyme inhibitors, AF – atrial fibrillation, ARB - angiotensin II receptor blockers, COPD– chronic obstructive pulmonary disease, CRT – cardiac resynchronization therapy, HF – heart failure, LBBB – left bundle branch block, LV – left ventricular, MRA – mineralocorticoid receptor antagonists, NYHA – New York Heart Association functional class, ROC – receiver operating characteristic

4.3.3.1 Top predictors of mortality in the 1- and 3-year cohorts

In the overall study population (including both sexes), the most important predictor of 1-year mortality was serum sodium, which was followed by serum creatinine, hemoglobin concentration, age, and etiology of HF (Figure 23). These features were also found among the strongest predictors of 3-year mortality, however, in different order of importance (serum sodium, age at implantation, hemoglobin concentration, serum creatinine, and etiology). Digitalis and the type of AF were found to show the most prominent change in their importance from 1 to 3 years (both $p < 0.001$).

4.3.3.2 Sex-specific patterns of mortality predictors at 1-year follow-up

We observed several sex-specific differences during the subgroup analysis. In males, the top predictors of 1-year mortality were hemoglobin concentration, serum sodium, serum creatinine, LBBB morphology, and age, whereas, in females, the most important predictors were serum sodium, etiology, LVEF, age, and serum creatinine (Figure 23).

The comparison of predictors by sex at 1-year revealed that etiology ($p < 0.001$), LVEF ($p < 0.001$), and treatment with amiodarone ($p < 0.01$) were at least twice as important in females as in males. Moreover, age at implantation and NYHA functional class were also significantly more predictive for 1-year mortality in women compared to men (both $p < 0.001$). Whereas, in males, hemoglobin concentration, type of the implanted device, treatment with allopurinol had significantly higher predictive power than in females (all $p < 0.001$).

4.3.3.3 Sex-specific patterns of mortality predictors at 3-year follow-up

In males, the strongest determinants of 3-year mortality were serum sodium, hemoglobin concentration, age at implantation, serum creatinine, and allopurinol, whereas, in females, these features were serum sodium, age at implantation, type of AF, NYHA functional class, and etiology in decreasing order (Figure 23).

Regarding females, NYHA functional class, etiology, LVEF, and type of AF exhibited significantly higher predictive power than in men (all $p < 0.001$). In males, features with at least a 2-fold higher importance were loop diuretics ($p < 0.001$), hemoglobin concentration ($p = 0.021$), allopurinol ($p < 0.001$), diabetes ($p < 0.001$), LV lead position ($p < 0.001$) and LBBB morphology ($p < 0.001$).

4.3.3.4 Longitudinal changes in the sex-specific patterns of predictors

We also identified features with the most notable changes in importance from 1 to 3 years of follow-up.

Among males, the most prominent increase of feature importance occurred in LV lead position, NYHA class, age, type of AF, hypertension, and digitalis (all $p < 0.001$). The importance of serum creatinine declined significantly ($p = 0.026$).

In females, we observed the greatest increase in the importance of NYHA functional class ($p < 0.001$), type of AF ($p < 0.001$), hypertension ($p < 0.001$), and age at implantation ($p < 0.014$). Among the top 10 predictors, the most considerable decrease from 1- to 3-year in feature importance was noted in the following factors: serum creatinine, LV end-diastolic diameter, QRS morphology, and amiodarone (all $p < 0.001$).

4.3.3.5 In-depth analysis of the associations between predictors and outcomes

The association between the most important predictors and the predicted outcome is visually presented in Figures 24 and 25. Older age, higher serum levels of creatinine, lower values of LVEF, serum sodium, hemoglobin concentration, ischemic etiology, non-LBBB morphology, higher NYHA classes, and the history of or current paroxysmal, persistent or permanent AF were associated with a higher predicted probability of 1- and 3-year all-cause mortality. Males exhibited higher values of predicted probability of all-cause death in all examined features compared to females. However, as ML models capture complex, high-level interactions among a multitude of variables, it is challenging to determine the effect of a single feature on the predicted probability of mortality, and the results of univariable analyses should be interpreted with caution.

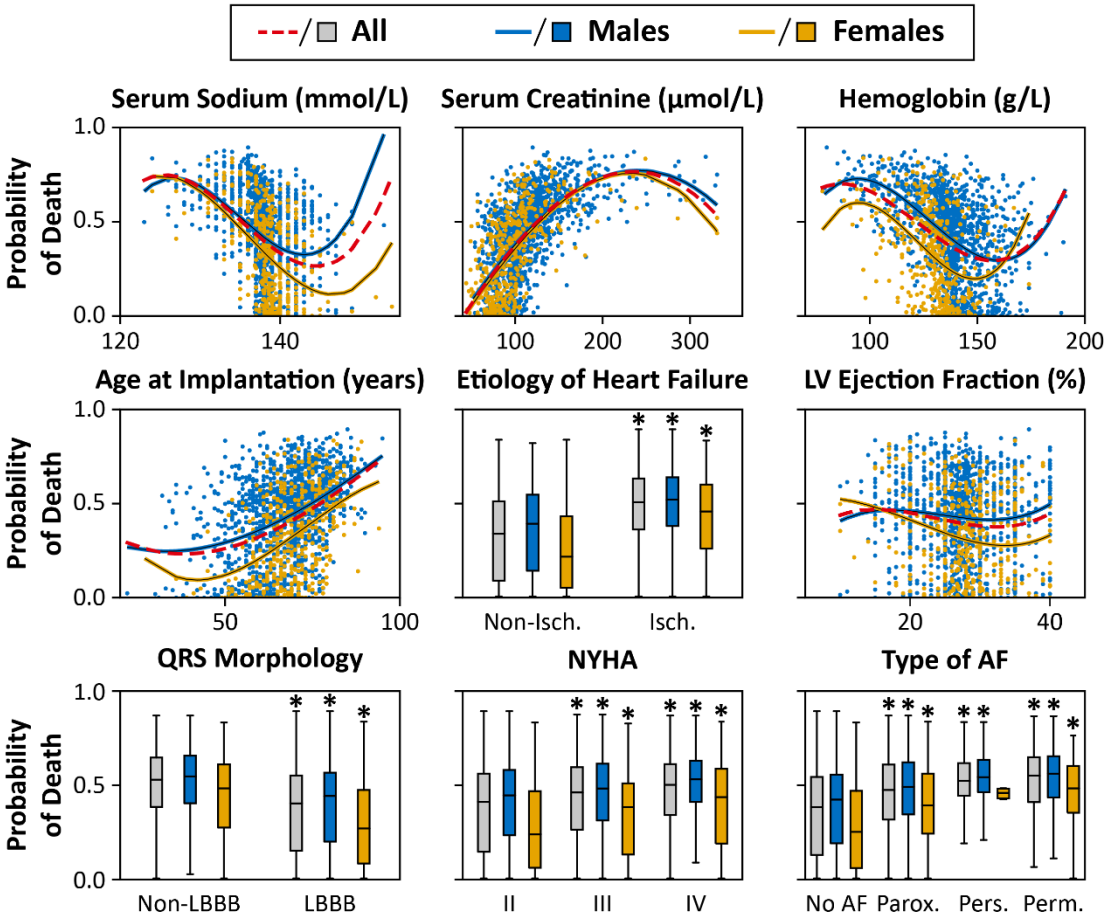


Figure 24 *Effect of the most important features on the predicted probability of 1-year all-cause mortality in the training cohorts (173)*

*The probability of death was calculated for each patient in the training cohort with 10-fold cross-validation. The predicted probability is plotted for each patient, and second-order polynomial trendlines are fitted to their values. *p < 0.05 vs. non-ischemic/non-LBBB morphology/NYHA class II/no AF, unpaired Student's t-test or Mann-Whitney U test.*

AF – atrial fibrillation, LBBB – left bundle branch block, LV – left ventricular, NYHA – New York Heart Association

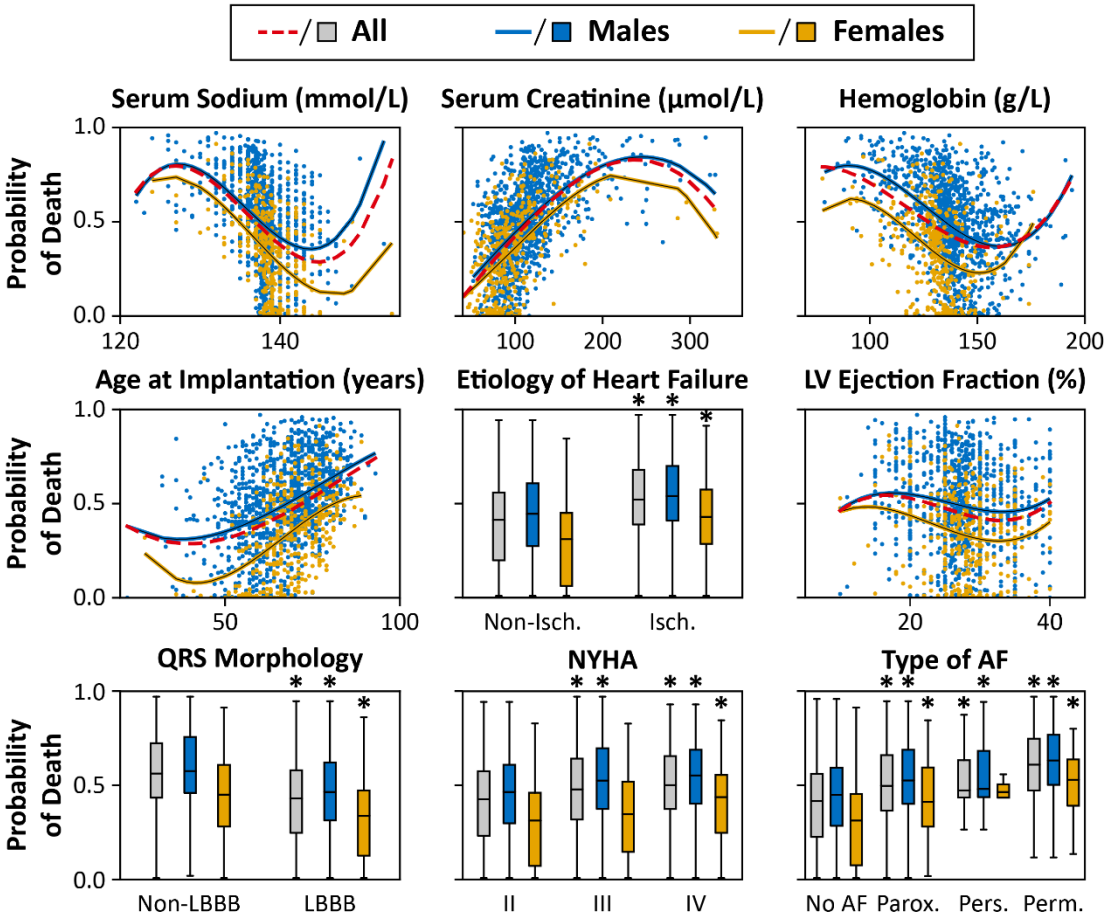


Figure 25 *Effect of the most important features on the predicted probability of 3-year all-cause mortality in the training cohorts (173)*

*The probability of death was calculated for each patient in the training cohort with 10-fold cross-validation. The predicted probabilities are plotted for each patient, and second-order polynomial trendlines are fitted to their values. *p < 0.05 vs. non-ischemic/non-LBBB morphology/ NYHA class II/no AF, unpaired Student's t-test or Mann-Whitney U test.*

AF – atrial fibrillation, LBBB – left bundle branch block, LV – left ventricular, NYHA – New York Heart Association

5. DISCUSSION

5.1 Applying topological data analysis to integrate echocardiographic features of left ventricular structure and function into a patient similarity network

The notion of patient similarity is a growing idea in personalized predictive analytics to support clinical assessment (157, 174-177). Patient similarity is a method that can empower precision medicine to stratify patients into clinically relevant subgroups (174). Such subgroup identification generally involves the use of unsupervised ML methods for clustering patients (15). However, most clustering techniques discretize the continuous patient data to develop discrete groups, using arbitrary thresholds. In contrast, TDA refers to a collection of powerful geometric approaches that integrate complex high-dimensional data to develop a patient-patient similarity network. TDA involves the generation of partially overlapping clusters and illustrates the entire study population as a continuous network of similar patients. This method can allow us to capture the notion of connectivity and continuum to describe the different stages of a disease.

Using retrospectively and prospectively collected echocardiographic data from 1,334 patients, we illustrated, for the first time, the potential role of a patient-patient similarity network for mapping cardiac dysfunction without the constraint of any a priori diagnostic system in varying degrees of LV structural and functional remodeling. Specifically, the TDA model in our analysis clustered the multi-parametric data without using a hierarchical structure or a branching tree but rather meaningfully represented the geometry of the data based on the similarity of the patients (50). Remarkably, the nodes clustered to produce a network in the form of a loop. Moreover, this loop demonstrated the relationships with the outcome of interest, which suggested a valid method of risk stratification for patients. We further illustrated the potential value of this loop for individualized predictions. Using a group of patients with longitudinally collected echocardiographic studies in which the patients were sampled in different stages of cardiac dysfunction, our analysis suggested that this looped space might also represent the periodic or recurrent behavior of the disease, thereby tracing the path that patients traveled through cycles of worsening cardiac function and recovery.

5.1.1 Patient similarity vs. an average patient

Echocardiography remains the most versatile tool in clinical practice, offering an ever-increasing array of measurements. Although novel multivariable data-driven analytic approaches to stratifying cardiac dysfunction have been recently made available (178), the integration of clinical and echocardiographic data for precision phenotyping has been arduous using traditional techniques.

Classically, studies and results are based on meticulous experimental designs and statistical analyses to produce exhaustive results for an average patient. However, no two patients are alike, making it difficult to generalize study results for average patients to actual patients with cardiac dysfunction. To this end, novel bioinformatics and machine-learning approaches have been suggested to support the integration of high-dimensional data for rapid medical decision-making (14, 179, 180).

The concept of the patient similarity network using TDA has been shown in well-known studies, such as those aiming to subgroup patients with diabetes (157), to identify individuals being resistant to malaria infections (160), and to uncover distinct phenotypes of aortic stenosis severity (181, 182). TDA has specifically enabled real-time exploration of the concept of disease space. For example, Torres et al. demonstrated a similar loop in analyzing disease tolerance to malaria in mice and humans and stratified the resilience of patients based on the size of the loop through the disease space (160). Similarly, we demonstrated a gradual change in echocardiographic variables throughout the disease cycle that outlined similarity among patients and described different phenotypes of cardiac functions in the disease space. Despite the abundance and complexity of echocardiographic features, distinct paths emerged for the patients with cardiac dysfunction in clockwise or counterclockwise directions on the loop based on the progression, treatment, and recovery of the disease.

5.1.2 Clinical implications

There are several pathophysiological and clinical implications for our study. First, the continuity of our patient similarity network suggested that the pathophysiological classification of cardiac dysfunction should be viewed as a continuum rather than as arbitrary divisions of the patient population into discrete subgroups as HF with reduced, mid-range, or preserved EF. Measures of LV systolic and diastolic function did not

exhibit abrupt changes at any level of cardiac function, but they covered a gradual and continuous spectrum, creating an overlapping and interconnected spectrum of disease phenotypes that was previously suggested but not shown (183). Second, the 4 regions of the loop showed incremental value over NYHA functional class, ACC/AHA stages, and commonly used risk scores (e.g., MAGGIC risk score), which suggested the clinical usefulness of this approach in patient risk stratification. Finally, unlike the consensus-driven algorithms (e.g., NYHA functional class and ACC/AHA stage) that first use expert knowledge and then develop the stages and decision pathways, the computational technique described in this study learns automatically and requires no a priori knowledge or training to develop meaningful disease representation. This ability to integrate multiple parameters pragmatically to define patient phenotypes and reproduce known clinical knowledge provides a strong foundation for why a provider could rely on this simplified staging scheme. Moreover, the TDA characterization system enables the identification of patients on a disease map much like the Global Positioning System; thus, it can facilitate automated classification, risk stratification, prognostication, or monitoring response to therapies. Such decision support systems are critically needed not only for clinical care but also for clinical trials in which heterogeneity of disease presentation affects patient matching and the discovery of novel therapies.

5.1.3 Validation in additional cohorts

Since the publication of this study (154), tremendous effort has been made to validate the prognostic power of our topological network in additional cohorts (184). To this end, we assigned each patient a low-risk (regions I and II) or high-risk (regions III and IV) label depending on their location on the TDA network loop structure, and we trained a deep neural network to predict the high- and low-risk phenogroups (184). In two external cohorts, the deep neural network model showed higher AUC than the 2016 American Society of Echocardiography guideline grades for predicting elevated LV filling pressure, and the high-risk phenogroup showed higher rates of HF hospitalization and/or death compared to the low-risk phenogroup (184). Similarly, in the TOPCAT trial cohort (90), the high-risk phenogroup showed higher rates of HF hospitalization or cardiac death and higher event-free survival with spironolactone therapy. Lastly, in the pooled RELAX-HF (Phosphodiesterase-5 Inhibition to Improve Clinical Status and

Exercise Capacity in Diastolic Heart Failure) /NEAT-HFpEF (Nitrate's Effect on Activity Tolerance in Heart Failure With Preserved Ejection Fraction) cohort (185, 186), the high-risk phenogroup had a higher burden of chronic myocardial injury, neurohormonal activation, and lower exercise capacity than the low-risk phenogroup (184). These findings imply that our topological network can be applied to reliably prognosticate patients with varying degrees of systolic and diastolic dysfunction.

5.2 Designing and evaluating a machine learning-based risk stratification system to predict all-cause mortality of patients undergoing CRT implantation

In our second study, we developed and tested an ML-based risk stratification tool to predict all-cause mortality of CRT patients during a 5-year follow-up period. Among the evaluated ML classifiers, random forest demonstrated the best performance; therefore, this algorithm was used to create the SEMMELWEIS-CRT score. With an average AUC over 0.700, the SEMMELWEIS-CRT score significantly outperformed the other currently available risk scores. We also developed an online calculator (available at semmelweiscrtscore.com) to enable a convenient, interactive, and personalized calculation of predicted mortality in patients undergoing CRT implantation.

5.2.1 Importance of risk assessment in patients undergoing CRT implantation

CRT induces reverse LV remodeling and improves outcomes in a certain subgroup of HF patients (143, 187). Despite these well-known beneficial effects, individual outcomes vary substantially. In the past years, several studies have investigated predictors that contribute to this variation, and numerous prognostic models have been developed by combining multiple risk factors (167-170). However, these currently available risk scores have shortcomings, and physicians are still reluctant to use them in daily clinical practice (125).

The major limitation is the insufficient reliability and ineffectiveness for risk assessment at the individual patient level, as outcome estimates have been extrapolated from large clinical trials. Although these scores offer general guidance and are effective at predicting outcomes at the population level, there remains a significant gap in the capability to predict outcomes for an individual patient (188). On the other hand,

individual prognostication remains essential to develop personalized treatment plans and make critical medical decisions based on life expectancy. These facts emphasize the need for more precise assessment through capturing the complex underlying interactions of predictors. With the SEMMELWEIS-CRT score, we intended to develop a more personalized approach for the risk assessment of patients undergoing CRT implantation.

5.2.2 Risk stratification with machine learning

Simultaneously interpreting the myriad risk predictors in an individual patient is challenging for clinicians. As a vast number of clinical variables associated with mortality is needed to be considered, the complexity of assessment increases, making it more difficult for clinicians to draw an overall conclusion regarding risk in an individual patient. Moreover, the potential influence of complex and hidden interactions between several weaker predictors is often overlooked. In this study, we demonstrated that ML is capable of overcoming these challenges by leveraging complex higher-level interactions among a multitude of clinical features. Accordingly, our model exhibited improved discrimination and predictive range with respect to all-cause mortality compared with the pre-existing risk scores. Moreover, the SEMMELWEIS-CRT score was capable of identifying patients with robustly increased risk of all-cause mortality (4th quartile) during the entire follow-up period.

With the increasing availability of enormous electronic data sets, ML algorithms have emerged as highly effective methods for medical prediction problems, with the potential to augment risk stratification (11). By making no a priori assumptions about causative factors, ML enables an agnostic exploration of all available data for non-linear patterns that may predict a particular individual's risk.

Our evaluation of ML algorithms was rigorous, including trials of numerous different classifiers within a wide hyper-parameter space. Among the evaluated algorithms, the best performing model was the random forest classifier which is consistent with previous studies using ML to predict clinical endpoints (127, 189-191).

There are various risk models available for the risk assessment of patients from the entire HF spectrum (171, 192). However, in our analysis, we focused exclusively on CRT recipients, and we generated models that recognize patterns in the clinical characteristics of this specific subset of HF patients. Moreover, many of the pre-existing

scores provide risk estimates for only a distinct time interval. In contrast, our goal was to build a model that could assess the risk of mortality annually from 1 to 5 years. Recently, Kalscheur et al. have developed an ML-based risk assessment tool, and their model exhibited comparable discriminative capabilities to ours (127). However, their model was limited to predicting 1-year outcomes, while the SEMMELWEIS-CRT score offers the prediction of mortality risk at 1-, 2-, 3-, 4-, and 5-year follow-up.

Ideally, ML models, such as the one developed in the present study, will be integrated into electronic medical record systems, and they will operate in the background providing real-time, personalized risk assessment based on the electronically available clinical features. Consequently, clinicians do not have to calculate a patient's risk manually, which may enhance the model's feasibility in clinical practice. Another potential benefit of ML algorithms is the capability to assimilate new data in real-time to continuously improve its own predictive accuracy.

The SEMMELWEIS-CRT score uses 33 clinical variables. The majority of them are routinely assessed during the management of HF; therefore, they are readily available from electronic medical records.

5.2.3 Future perspectives

The observed high efficacy of our random forest model suggests that ML should be integrated into the individual risk assessment of patients undergoing CRT implantation. We foresee that the role of ML-based prognostic risk scores will become increasingly relevant in the near future, and structured, dense databases in combination with state-of-the-art analytic approaches will pave the way to precision cardiovascular medicine.

5.3 Exploring the sex-specific differences and similarities in the predictors of mortality among patients undergoing CRT implantation

Using data from a single-center cohort of HF patients undergoing CRT implantation, we developed and evaluated ML-based algorithms for the prediction of 1- and 3-year all-cause mortality. The resulting CIRF models demonstrated good discriminatory power in assessing the risk of mortality with an AUC over 0.700 at 1- and 3-year follow-up. Moreover, ML performed substantially well across patient subsets containing exclusively males or females (AUCs ranging from 0.681 to 0.798). Serum sodium, creatinine, hemoglobin, age, and HF etiology were among the most important determinants of short- and mid-term mortality; however, their relative importance varied over time. As expected, female sex was associated with significantly better survival rates in our cohort as well. Sex-specific patterns were also identified in the predictors of mortality. The role of HF etiology (ischemic or non-ischemic), NYHA functional class, and LVEF were more pronounced in females, whereas hemoglobin concentration, QRS morphology, and treatment with allopurinol were notably more predictive for all-cause mortality in males.

5.3.1 Risk stratification of heart failure patients using machine learning

The personalized prediction of prognosis is fundamental to patient-centered care, both in optimizing treatment strategies and informing patients as part of shared decision-making. For this purpose, an abundance of prediction models has been developed; however, most of them had achieved only modest success, particularly when they were applied in HF populations other than those from which the scores were derived (125, 193). The unsatisfactory results of previous HF risk scores are likely due to multiple causes, including the fact that most of them were created using conventional statistical methods that failed to capture high-dimensional interactions among predictors that bear relevant prognostic information.

In contrast to traditional statistics, ML was explicitly designed to reveal and harness these correlations. Several studies have proved that these advanced data analytic approaches can leverage the complex, higher-level interplay between predictors and outcomes to achieve better discrimination. ML can improve the care of HF patients in

various ways, e.g., by augmenting the prediction of readmission after HF hospitalization or by predicting the risk of mortality (105, 108, 194). In HF patients undergoing CRT implantation, our research group has confirmed the superiority of ML over pre-existing risk scores in the previously presented study (166), and similar results have been reported by others as well (127, 129). Underpinning these findings, we were able to predict the 1- and 3-year mortality of CRT patients with good discrimination and excellent calibration, even in subsets of patients divided by sex. In light of the promising results of our single-center study, we will endeavor to validate our models in external cohorts in a multi-centric manner.

In our analysis, CIRF exhibited the best discriminative ability for predicting both 1- and 3-year mortality. To understand the outstanding performance of tree-based approaches such as CIRF in outcome prediction, an important difference between conventional regression models and tree-based methods should be highlighted. The former favors variables that have a uniform effect across the entire patient population, whereas the latter can uncover variables that might act differently in different patient subgroups. This is essential for personalized prognostication as in an individual patient, the discriminatory power of a given feature may be significantly enhanced or overshadowed by others. Due to this attribute, tree-based methods such as TRF and CIRF are extremely suitable for application as clinical decision-making tools (195).

Similar to TRF, CIRF is also an ensemble of individual decision trees; however, there is a major difference between the two algorithms: CIRF uses statistical theory (i.e., permutation test) based covariate selection scheme during the training process, whereas TRF creates its trees by selecting the feature that maximizes an information measure (e.g., Gini coefficient) at each split (30, 196). Accordingly, the method used in TRF tends to select features that have many possible splits (e.g., continuous or high-cardinality categorical variables); however, CIRF avoids this potential bias and enables the more appropriate calculation of feature importances.

5.3.2 Sex-specific differences in outcomes following CRT implantation

Sex is increasingly recognized as an important modulator of outcomes in CRT patients, and several studies such as the MADIT-CRT (149), the RAFT (Resynchronization-Defibrillation for Ambulatory Heart Failure Trial) (197), and the

MASCOT (Management of Atrial Fibrillation Suppression in AF-HF Comorbidity Therapy) (198) trials have suggested a greater CRT benefit in women. Despite the expanding knowledge about sex-related differences in HFrEF, the reason women benefit more than men from CRT remains unclear (153). Numerous plausible explanations have been proposed, such as the dissimilarities between sexes in the frequency of ischemic cardiomyopathy (199), AF, and comorbidities (148), or the sex-related differences in body height, LV size, and QRS duration (200, 201). In addition, the impact of sex hormones on the pathophysiology of HF or the sex-specific characteristics of pharmacodynamics and pharmacokinetics are also considerable factors (144, 202).

The sex-specific effects of QRS prolongation and morphology on outcomes have been intensively investigated in CRT patients (199, 201, 203-206). Thus, the findings of these studies have prompted calls for sex-specific guideline recommendations regarding the selection of CRT recipients. As women have shorter QRS durations than men in the absence of any conduction delay, they are more likely to exhibit a true LBBB compared to men at shorter QRS duration (207, 208). It has also been reported that among patients with LBBB and non-ischemic etiology, women have electrical dyssynchrony more frequently compared to men at any given QRS duration, and consequently, they would exhibit a better response to CRT (204). According to the study conducted by Beela et al., the interaction between HF etiology and mechanical dyssynchrony seems to represent another important aspect: due to the lower rate of ischemic etiology and the lower extent of scarred myocardium, women have more frequently uncomplicated patterns of LBBB-like mechanical dyssynchrony which is better amendable by CRT (199).

The beneficial effects of CRT also depend on device programming and the percentage of effective biventricular pacing. Notably, the latter significantly varies by sex, and therefore, sex-specific CRT programming has attracted increased attention (209). According to the results of the SMART-AV trial, the optimization of atrioventricular delay intervals is associated with improved outcomes in women but not in men (210), which might be attributable to the inherent sex-related differences in atrial geometry and PR intervals. A higher percentage of biventricular pacing has also been reported in women (198, 210, 211), most probably due to the lower rate of AF compared to men (212, 213). This could also contribute to the observed differences in mortality between sexes as even a small increment in the biventricular pacing rate may improve outcomes (214).

Although there are still many open questions, it is clear that multiple intercorrelated factors contribute to this phenomenon. Therefore, during the search for answers, ML-based approaches may come in handy, as they are particularly helpful in uncovering hidden patterns in large data sets by simultaneously interpreting predictors even in the presence of complex, non-linear interactions.

5.3.3 Sex-specific patterns in mortality predictors

Given the sex-related differences in the anatomy and physiology of the cardiovascular system, encountering dissimilarities in the importance of prognostic predictors between males and females is to be expected in CRT patients. Nevertheless, there is only a limited number of publications dedicated to the thorough exploration of this topic. To the best of our knowledge, our study is the first that evaluated the sex-related differences and similarities in mortality predictors of CRT patients using ML. In our analysis, we observed significant variations in the importance of several predictors such as HF etiology, NYHA functional class, LVEF, and AF between sexes, to name a few.

Utilizing the tools of conventional statistics, the sex-specific prognostic value of HF etiology has been previously investigated in large cohorts of HFrEF patients. In the MAGGIC meta-analysis, the ischemic etiology appeared to attenuate the protective effect of female sex on prognosis (215). In addition, ischemic cardiomyopathy and the extent of myocardial scar were found to be significant predictors of mortality in females but not in males among CRT patients (199). In line with this evidence, the paramount importance of HF etiology in women was proved in our study as well.

When analyzing the interaction between sex and different covariates in the prediction of survival after CRT implantation, Beela et al. reported that NYHA class was a significant predictor in males only (199). Moreover, among HFrEF patients, NYHA class had a more prominent prognostic value in men than in women (216). Contrary to these findings, a stronger association of NYHA functional class with outcomes was observed in females in our current analysis and the BEST (Beta-Blocker Evaluation in Survival Trial) as well (217).

Another well-established prognostic factor is LVEF, whose interaction with sex in the prediction of all-cause death has been demonstrated in CRT patients (199).

Complementing these findings and the results of the BEST trial (217), we have also demonstrated that LVEF is a stronger predictor of prognosis in women than in men.

In HFrEF patients, most studies agree on the prognostic value of AF; however, there is some inconsistency regarding its exact role as some investigations attribute more prognostic impact to AF in females (217), whereas others observed comparable predictive power in males and females (199, 216). Our results support the former as we found AF to have a more prominent effect on outcomes in females.

According to our analysis, the prognostic relevance of hyponatremia and renal function should also be emphasized in CRT patients. Our results are in accordance with the findings of Zusterzeel et al., who reported that despite being significant determinants in both sexes, serum creatinine and hyponatremia appeared to be stronger predictors in women than in men (203).

Lately, the interplay between sex and diabetes in HFrEF patients has attracted increased attention among researchers. Confirming the findings of the MAGGIC (215), the recently published analysis of the ASIAN-HF (Asian Sudden Cardiac Death in Heart Failure) registry demonstrated that diabetes is coupled with a greater risk of adverse outcomes in women than in men (218). In contrast, diabetes was associated with a higher risk of all-cause death or HF hospitalization in males in the Swedish Heart Failure Registry (216), and it was proven to be a significant predictor only in men in the BEST trial (217). Interestingly, in our study, diabetes was not ranked among the top five predictors in any of the analyzed patient subsets, and we detected inter-sex differences in its importance only at 3-year follow-up.

Some of our findings coincide with those of previous studies, whereas some others may not. These apparent discrepancies might be partly attributable to the fact that most studies applied Cox proportional hazards regression, whereas we utilized an entirely different methodology that captures other aspects of associations between risk factors and outcomes. Although the exact reasons behind these contradicting results should be clarified in further investigations, our findings underscore the necessity of sex-specific approaches in the management of HFrEF patients.

5.4 Limitations

Despite the highlighted strengths of the presented studies, each of them has limitations to be acknowledged.

In the first study, the follow-up duration and sample size for patients with reduced EF were modest and potentially averted us from capturing a greater number of cardiac events to test the applicability, disease trajectory over time, and ability of our model to predict measurable isolated endpoints. Furthermore, specific therapeutic interventions that targeted any specific region on the loop or types of therapies that could change the patients' prognoses were not included in the study. This represents a logical next step that would be required to address in future studies. The addition of biomarkers and novel echocardiographic parameters, such as strain and strain rate, could provide an added benefit to the model and should also be investigated in the future.

As the second and third studies have similar design and methodology, they have some limitations in common. First, both studies represent results from a single center. As we were aware of this limitation, we performed hyperparameter tuning with 10-fold cross-validation in the training cohorts, and we also tested our models in statistically independent test cohorts to enhance generalizability. Nonetheless, as the next step, the robustness of our models should be tested in external populations as well. Second, the utilized databases bear the inherent limitations of retrospective data collection, such as the higher proportion of missing data (compared to prospective trials) and the heterogeneity partly attributable to the changes in guideline recommendations over the years. However, the use of such real-world data holds the potential for better generalizability. Third, our models use baseline (pre-implantation and procedural) variables without incorporating the time-varying values of these parameters. Although a dynamic model integrating values of the same parameter from multiple time points may be superior, in the present studies, we aimed to predict all-cause mortality using clinical data that could be acquired at the time of device implantation. Finally, there may remain additional domains of variables (e.g., imaging data, novel biomarkers, genetics, or quality of life questionnaires) that could further improve the predictive capability of our models. Future work should explore the addition of such features to enhance the models proposed in the presented studies.

6. CONCLUSIONS

1. Based on our first study, in which we applied TDA to analyze the retrospectively and prospectively collected echocardiographic data of a large patient cohort with varying degrees of LV structural and functional remodeling, we reached the following conclusions:
   1.1 TDA is a robust data analytical approach that is capable of effectively integrating multiple echocardiographic parameters of LV structure and function into a looped patient-patient similarity network in which subjects could be mapped to specific locations associated with distinct disease stages and clinical outcomes.
   1.2 TDA can be utilized to trace the progression of cardiac dysfunction in patients as they travel through cycles of compensation and decompensation within a looped disease space.
   1.3 TDA may have broad implications for developing clinical risk stratification schemes using patient-patient similarity networks and bring us one step closer to precision medicine.
2. In our second study, in which we implemented and evaluated various ML algorithms to predict all-cause mortality in patients undergoing CRT implantation, we came to the following conclusions:
   2.1 By capturing the non-linear associations between predictors and outcomes, our CIRF-based risk stratification system – the SEMMELWEIS-CRT score – achieved high performance in predicting 1-, 2-, 3-, 4-, and 5-year all-cause death in patients undergoing CRT implantation and effectively outlined patient subgroups at high risk for mid- and long-term mortality.
   2.2 The SEMMELWEIS-CRT score outperformed several conventional statistics-based risk scores, namely the SHFM, the VALID-CRT, the EAARN, the ScREEN, and the CRT-score.
   2.3 The proposed model laid the foundation stone for future testing of its clinical utility as a decision support tool to optimize candidate selection and improve the prognostication of CRT patients.
3. Our third study, in which we used ML algorithms to explore the sex-specific differences and similarities in the predictors of all-cause mortality among patients undergoing CRT, led us to the following conclusions:

3.1 Female sex was found to be associated with significantly better survival rates in our cohort of CRT patients.

3.2 Using CIRF in combination with easily obtainable clinical features, we could effectively predict 1- and 3-year all-cause mortality in patient subsets containing males or females exclusively.

3.3 Sex-specific patterns were identified in the predictors of mortality, which also changed over time. The role of HF etiology (ischemic or non-ischemic), NYHA functional class, and LVEF were more pronounced in females, whereas hemoglobin concentration, QRS morphology, and treatment with allopurinol were notably more predictive for all-cause mortality in males.

7. SUMMARY

In cardiovascular medicine, one of the major priorities is to prevent adverse clinical events and hospitalization by risk factor management and by earlier detection of subclinical cardiac dysfunction. To improve predictive modeling and elucidate novel determinants of a specific outcome, ML has been increasingly applied in HF research. ML algorithms can account for interactions between myriads of predictors and their non-linear associations with outcomes; thus, their utilization could potentially lead to improved predictive models. In this thesis, the potentials of ML were demonstrated through three studies aiming to improve the prognostication of HF patients.

Using retrospectively and prospectively collected echocardiographic data, we illustrated that a patient-patient similarity network is suitable for mapping cardiac dysfunction without the constraint of any a priori diagnostic system in varying degrees of LV structural and functional remodeling. By integrating echocardiographic features, TDA created a looped network in which patients could be mapped to specific locations associated with distinct disease stages and clinical outcomes. Moreover, the created topological network enabled us to trace the progression of cardiac dysfunction in patients as they travel through cycles of worsening cardiac function and recovery within a looped disease space.

We also demonstrated that ML can capture high-dimensional, non-linear relationships among clinical features of patients undergoing CRT implantation, and it can be used to develop a risk stratification system that predicts 1-, 2-, 3-, 4-, and 5-year all-cause mortality more accurately than the currently available, conventional statistics-based risk scores. As the relative importance of features explaining sex-related differences in outcomes is scarcely explored in this patient population, we also aimed to assess the sex-specific differences and similarities in predictors of 1- and 3-year all-cause mortality utilizing ML techniques. Our analysis revealed sex-specific patterns in the predictors of mortality, which also changed over time.

In summary, ML and TDA offer practical and robust solutions for personalized risk stratification and prognostication in cardiovascular medicine and may enable the identification of patient populations who are more likely to respond to a specific therapy.

8. ÖSSZEFOGLALÁS

A kardiovaszkuláris medicinában kiemelt jelentőséggel bír a nemkívánatos klinikai események hatékony megelőzése, melyhez elengedhetetlen a rizikófaktorok megfelelő kezelése és a szubklinikus kardiális diszfunkció korai felismerése. Ezt a célt szolgáló prediktív modellek tökéletesítése és az adott klinikai végpont prediktorainak feltárása érdekében egyre szélesebb körben kerül alkalmazásra a gépi tanulás. A gépi tanulásban rejlő lehetőségeket három olyan vizsgálaton keresztül mutattuk be, amelyek célja a szívelégtelenségben szenvedő betegek prognosztikációjának javítása.

Retrospektív és prospektív echokardiográfiás adatbázisok felhasználásával bemutattuk, hogy különböző mértékű strukturális és funkcionális bal kamra remodellációt mutató betegek úgynevezett hasonlósági hálózata alkalmas a kardiális diszfunkció teljes spektrumának feltérképezésére. Az echokardiográfiás paraméterek topologikus elemzésével egy olyan körkörös hálózatot hoztunk létre, melynek egyes régiói a kardiális diszfunkció bizonyos stádiumainak feleltek meg, illetve eltérést mutattak a klinikai kimenetel tekintetében is. Továbbá, a topológiai hálózat segítségével lehetőségünk volt a kardiális diszfunkció progressziójának nyomon követesére az egyes betegeknél.

Szintén demonstráltuk, hogy a gépi tanulás képes modellezni a kardiális reszinkronizációs terápián áteső betegek klinikai jellemzőiben rejlő sokdimenziós, nemlineáris kapcsolatokat. Így ezek az algoritmusok alkalmasak voltak egy olyan rizikó stratifikációs rendszer létrehozására, amely pontosabban prediktálta az 1, 2, 3, 4 és 5 éves összhalálozást, mint a jelenleg rendelkezésre álló, hagyományos statisztikán alapuló rizikóbecslő pontrendszerek. Mivel az összmortalitás prediktorainak relatív fontosságában fellelhető nemi különbségeket korábban alig vizsgálták ebben a betegpopulációban, ezért gépi tanulási technikák segítségével meghatároztuk az 1- és 3 éves összhalálozás prediktorainak nemek közti különbségeit. Elemzésünk a nők és férfiak esetében is karakterisztikus mintázatokat tárt fel az összmortalitás prediktoraiban, amelyek az implantációtól eltelt idő függvényében jelentős változásokat mutattak.

Összefoglalva, a gépi tanulás és a topologikus adatelemzés robusztus megoldást kínál a kardiovaszkuláris betegek személyre szabott rizikóbecslésére és prognosztikációjára, illetve lehetőséget nyújt azon betegpopulációk azonosítására, amelyek nagyobb valószínűséggel reagálnak kedvezően egy adott kezelésre.

9. REFERENCES

1. Obermeyer Z, Lee TH. (2017) Lost in Thought — The Limits of the Human Mind and the Future of Medicine. New England Journal of Medicine, 377: 1209-1211.
2. Deo RC. (2015) Machine Learning in Medicine. Circulation, 132: 1920-1930.
3. Rajkomar A, Dean J, Kohane I. (2019) Machine Learning in Medicine. New England Journal of Medicine, 380: 1347-1358.
4. Turing AM. (1950) I. Computing Machinery and Intelligence. Mind, LIX: 433-460.
5. Minsky M. Theory of Neural-Analog Reinforcement System and its Applications to the Brain-Model Problem. Vol. Ph.D. University of Princeton, 1954.
6. McCarthy J, Minsky M, Shannon CE, Rochester N, Dartmouth C. (1955) A proposal for the Dartmouth Summer Research Project on Artificial Intelligence.
7. McCarthy J, Minsky ML, Rochester N, Shannon CE. (2006) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine, 27: 12.
8. Samuel AL. (1959) Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3: 210-229.
9. Langley P. (2011) The changing science of machine learning. Machine Learning, 82: 275-279.
10. Quer G, Arnaout R, Henne M, Arnaout R. (2021) Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. Journal of the American College of Cardiology, 77: 300-313.
11. Al'Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, Pandey M, Maliakal G, van Rosendael AR, Beecy AN, Berman DS, Leipsic J, Nieman K, Andreini D, Pontone G, Schoepf UJ, Shaw LJ, Chang HJ, Narula J, Bax JJ, Guan Y, Min JK. (2019) Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J, 40: 1975-1986.
12. Johnson KW, Soto JT, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E, Dudley JT. (2018) Artificial Intelligence in Cardiology. Journal of the American College of Cardiology, 71: 2668-2679.

13. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, Marwick TH. (2019) Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. J Am Coll Cardiol, 73: 1317-1335.
14. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. (2017) Artificial Intelligence in Precision Cardiovascular Medicine. J Am Coll Cardiol, 69: 2657-2664.
15. Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. (2019) Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. J Am Heart Assoc, 8: e012788.
16. Goldstein BA, Navar AM, Carter RE. (2017) Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J, 38: 1805-1814.
17. Bazoukis G, Stavrakis S, Zhou J, Bollepalli SC, Tse G, Zhang Q, Singh JP, Armoundas AA. (2021) Machine learning versus conventional clinical methods in guiding management of heart failure patients-a systematic review. Heart Fail Rev, 26: 23-34.
18. Gandomi A, Haider M. (2015) Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35: 137-144.
19. White IR, Royston P, Wood AM. (2011) Multiple imputation using chained equations: Issues and guidance for practice. Stat Med, 30: 377-399.
20. Hayati Rezvan P, Lee KJ, Simpson JA. (2015) The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. BMC Medical Research Methodology, 15: 30.
21. Rudin C. (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1: 206-215.
22. Goldstein BA, Pencina MJ, Montez-Rath ME, Winkelmayer WC. (2017) Predicting mortality over different time horizons: which data elements are needed? J Am Med Inform Assoc, 24: 176-181.
23. Hoerl AE, Kennard RW. (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12: 55-67.

24. Hoerl AE, Kennard RW. (1970) Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, 12: 69-82.
25. Tibshirani R. (1996) Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58: 267-288.
26. Zou H, Hastie T. (2005) Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67: 301-320.
27. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. Taylor & Francis1984.
28. Loh W-Y. (2011) Classification and regression trees. WIREs Data Mining and Knowledge Discovery, 1: 14-23.
29. Breiman L. (1996) Bagging predictors. Machine Learning, 24: 123-140.
30. Breiman L. (2001) Random Forests. Machine Learning, 45: 5-32.
31. Freund Y, Schapire RE. (1997) A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55: 119-139.
32. Friedman JH. (2001) Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29: 1189-1232, 1144.
33. Cortes C, Vapnik V. (1995) Support-vector networks. Machine Learning, 20: 273-297.
34. Cover T, Hart P. (1967) Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13: 21-27.
35. Webb GI. Bayes Rule. In: Sammut C, Webb GI (szerk.), Encyclopedia of Machine Learning, doi:10.1007/978-0-387-30164-8_62. Springer US, Boston, MA, 2010: 74-75.
36. Hand DJ, Yu K. (2001) Idiot's Bayes: Not So Stupid after All? International Statistical Review / Revue Internationale de Statistique, 69: 385-398.
37. Frank E, Trigg L, Holmes G, Witten IH. (2000) Technical Note: Naive Bayes for Regression. Machine Learning, 41: 5-25.
38. McCallum A, Nigam K. A Comparison of Event Models for Naive Bayes Text Classification. In: Learning for Text Categorization: Papers from the 1998 AAAI Workshop, 1998: 41--48.

39. Lloyd S. (1982) Least squares quantization in PCM. IEEE Transactions on Information Theory, 28: 129-137.
40. Reynolds D. Gaussian Mixture Models. In: Li SZ, Jain A (szerk.), Encyclopedia of Biometrics, doi:10.1007/978-0-387-73003-5_196. Springer US, Boston, MA, 2009: 659-663.
41. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D. (2016) Mastering the game of Go with deep neural networks and tree search. Nature, 529: 484-489.
42. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D. (2015) Human-level control through deep reinforcement learning. Nature, 518: 529-533.
43. Kober J, Bagnell JA, Peters J. (2013) Reinforcement learning in robotics: A survey. The International Journal of Robotics Research, 32: 1238-1274.
44. Tai L, Liu M. (2016) Mobile robots exploration through cnn-based reinforcement learning. Robotics Biomim, 3: 24.
45. Sallab A, Abdou M, Perot E, Yogamani S. (2017) Deep Reinforcement Learning framework for Autonomous Driving. Electronic Imaging, 2017: 70-76.
46. Chopra R, Roy SS. End-to-End Reinforcement Learning for Self-driving Car. In: Advanced Computing and Intelligent Engineering (Pati B, Panigrahi CR, Buyya R, Li K-C szerk.). Springer Singapore, Singapore, 2020: 53-61.
47. Yu C, Liu J, Zhao H. (2019) Inverse reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. BMC Medical Informatics and Decision Making, 19: 57.
48. Levy AE, Biswas M, Weber R, Tarakji K, Chung M, Noseworthy PA, Newton-Cheh C, Rosenberg MA. (2020) Applications of machine learning in decision analysis for dose management for dofetilide. PLOS ONE, 14: e0227324.
49. Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U, Min JK, Tang WHW, Halperin JL, Narayan SM. (2019) Deep learning for cardiovascular medicine: a practical primer. Eur Heart J, 40: 2058-2073.

50. Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G. (2013) Extracting insights from the shape of complex data using topology. Scientific Reports, 3: 1236.
51. Carlsson G. (2009) Topology and data. Bulletin of the American Mathematical Society, 46: 255-308.
52. Singh G, Memoli F, Carlsson G. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In: Eurographics Symposium on Point-Based Graphics, doi:10.2312/SPBG/SPBG07/091-100. The Eurographics Association, 2007.
53. Drummond C, Holte RC. (2006) Cost curves: An improved method for visualizing classifier performance. Machine Learning, 65: 95-130.
54. Bozkurt B, Coats AJS, Tsutsui H, Abdelhamid CM, Adamopoulos S, Albert N, Anker SD, Atherton J, Böhm M, Butler J, Drazner MH, Michael Felker G, Filippatos G, Fiuzat M, Fonarow GC, Gomez-Mesa JE, Heidenreich P, Imamura T, Jankowska EA, Januzzi J, Khazanie P, Kinugawa K, Lam CSP, Matsue Y, Metra M, Ohtani T, Francesco Piepoli M, Ponikowski P, Rosano GMC, Sakata Y, Seferović P, Starling RC, Teerlink JR, Vardeny O, Yamamoto K, Yancy C, Zhang J, Zieroth S. (2021) Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure: Endorsed by the Canadian Heart Failure Society, Heart Failure Association of India, Cardiac Society of Australia and New Zealand, and Chinese Heart Failure Association. Eur J Heart Fail, 23: 352-380.
55. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. (2020) Epidemiology of heart failure. European Journal of Heart Failure, 22: 1342-1356.
56. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, Burri H, Butler J, Čelutkienė J, Chioncel O, Cleland JGF, Coats AJS, Crespo-Leiro MG, Farmakis D, Gilard M, Heymans S, Hoes AW, Jaarsma T, Jankowska EA, Lainscak M, Lam CSP, Lyon AR, McMurray JJV, Mebazaa A, Mindham R, Muneretto C, Francesco Piepoli M, Price S, Rosano GMC, Ruschitzka F, Kathrine Skibelund A, Group ESD. (2021) 2021 ESC Guidelines for the diagnosis and

treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. European Heart Journal, 42: 3599-3726.

57. Glikson M, Nielsen JC, Kronborg MB, Michowitz Y, Auricchio A, Barbash IM, Barrabés JA, Boriani G, Braunschweig F, Brignole M, Burri H, Coats AJS, Deharo J-C, Delgado V, Diller G-P, Israel CW, Keren A, Knops RE, Kotecha D, Leclercq C, Merkely B, Starck C, Thylén I, Tolosana JM, Group ESD. (2021) 2021 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: Developed by the Task Force on cardiac pacing and cardiac resynchronization therapy of the European Society of Cardiology (ESC) With the special contribution of the European Heart Rhythm Association (EHRA). European Heart Journal, 42: 3427-3520.
58. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA. (2020) Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. American Heart Journal, 229: 1-17.
59. Karanasiou GS, Tripoliti EE, Papadopoulos TG, Kalatzis FG, Goletsis Y, Naka KK, Bechlioulis A, Errachid A, Fotiadis DI. (2016) Predicting adherence of patients with HF through machine learning techniques. Healthcare Technology Letters, 3: 165-170.
60. Son YJ, Kim HG, Kim EH, Choi S, Lee SK. (2010) Application of support vector machine for prediction of medication adherence in heart failure patients. Healthc Inform Res, 16: 253-259.
61. Graven LJ, Higgins MK, Reilly CM, Dunbar SB. (2020) Heart Failure Symptoms Profile Associated With Depressive Symptoms. Clinical Nursing Research, 29: 73-83.
62. Dini FL, Ballo P, Badano L, Barbier P, Chella P, Conti U, De Tommasi SM, Galderisi M, Ghio S, Magagnini E, Pieroni A, Rossi A, Rusconi C, Temporelli PL. (2010) Validation of an echo-Doppler decision model to predict left ventricular filling pressure in patients with heart failure independently of ejection fraction. European Journal of Echocardiography, 11: 703-710.

63. Hei X, Du X, Wu J, Hu F. Defending Resource Depletion Attacks on Implantable Medical Devices. In: 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, doi:10.1109/GLOCOM.2010.5685228, 2010: 1-5.
64. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. (2019) Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med, 25: 70-74.
65. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, Carter RE, Shah ND, Friedman PA, Noseworthy PA. (2019) Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. J Cardiovasc Electrophysiol, 30: 668-674.
66. Adedinsewo D, Carter RE, Attia Z, Johnson P, Kashou AH, Dugan JL, Albus M, Sheele JM, Bellolio F, Friedman PA, Lopez-Jimenez F, Noseworthy PA. (2020) Artificial Intelligence-Enabled ECG Algorithm to Identify Patients With Left Ventricular Systolic Dysfunction Presenting to the Emergency Department With Dyspnea. Circ Arrhythm Electrophysiol, 13: e008437.
67. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM, Song PS, Park J, Choi RK, Oh BH. (2019) Development and Validation of Deep-Learning Algorithm for Electrocardiography-Based Heart Failure Identification. Korean Circ J, 49: 629-639.
68. Kagiyama N, Piccirilli M, Yanamala N, Shrestha S, Farjo PD, Casaclang-Verzosa G, Tarhuni WM, Nezarat N, Budoff MJ, Narula J, Sengupta PP. (2020) Machine Learning Assessment of Left Ventricular Diastolic Function Based on Electrocardiographic Features. J Am Coll Cardiol, 76: 930-941.
69. Kwon J-m, Kim K-H, Eisen HJ, Cho Y, Jeon K-H, Lee SY, Park J, Oh B-H. (2020) Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features. European Heart Journal - Digital Health, 2: 106-116.
70. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. (2018) Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation, 138: 1623-1635.

71. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY. (2020) Video-based AI for beat-to-beat assessment of cardiac function. Nature, 580: 252-256.
72. Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, Romano N, Hong H, Mor-Avi V, Martin RP, Lang RM. (2019) Automated Echocardiographic Quantification of Left Ventricular Ejection Fraction Without Volume Measurements Using a Machine Learning Algorithm Mimicking a Human Expert. Circ Cardiovasc Imaging, 12: e009303.
73. Reddy YNV, Carter RE, Obokata M, Redfield MM, Borlaug BA. (2018) A Simple, Evidence-Based Approach to Help Guide Diagnosis of Heart Failure With Preserved Ejection Fraction. Circulation, 138: 861-870.
74. Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Marino P, Fraser AG, D'Hooge J. (2018) Diagnosis of Heart Failure With Preserved Ejection Fraction: Machine Learning of Spatiotemporal Variations in Left Ventricular Deformation. J Am Soc Echocardiogr, 31: 1272-1284.e1279.
75. Sanchez-Martinez S, Duchateau N, Erdei T, Kunszt G, Aakhus S, Degiovanni A, Marino P, Carluccio E, Piella G, Fraser AG, Bijnens BH. (2018) Machine Learning Analysis of Left Ventricular Function to Characterize Heart Failure With Preserved Ejection Fraction. Circulation: Cardiovascular Imaging, 11: e007138.
76. Blecker S, Katz SD, Horwitz LI, Kuperman G, Park H, Gold A, Sontag D. (2016) Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data. JAMA Cardiology, 1: 1014-1020.
77. Blecker S, Sontag D, Horwitz LI, Kuperman G, Park H, Reyentovich A, Katz SD. (2018) Early Identification of Patients With Acute Decompensated Heart Failure. J Card Fail, 24: 357-362.
78. Zhang R, Ma S, Shanahan L, Munroe J, Horn S, Speedie S. (2018) Discovering and identifying New York heart association classification from electronic health records. BMC Med Inform Decis Mak, 18: 48.
79. Meystre SM, Kim Y, Gobbel GT, Matheny ME, Redd A, Bray BE, Garvin JH. (2016) Congestive heart failure information extraction framework for automated

treatment performance measures assessment. Journal of the American Medical Informatics Association, 24: e40-e46.

80. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, Gomes AS, Folsom AR, Shea S, Guallar E, Bluemke DA, Lima JAC. (2017) Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. Circ Res, 121: 1092-1101.
81. Chahal H, Bluemke DA, Wu CO, McClelland R, Liu K, Shea SJ, Burke G, Balfour P, Herrington D, Shi P, Post W, Olson J, Watson KE, Folsom AR, Lima JA. (2015) Heart failure risk prediction in the Multi-Ethnic Study of Atherosclerosis. Heart, 101: 58-64.
82. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, Basit M, Kannan V, Grodin JL, Everett B, Willett D, Berry J, Pandey A. (2019) Machine Learning to Predict the Risk of Incident Heart Failure Hospitalization Among Patients With Diabetes: The WATCH-DM Risk Score. Diabetes Care, 42: 2298-2306.
83. Choi E, Schuetz A, Stewart WF, Sun J. (2017) Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc, 24: 361-370.
84. Chen R, Stewart WF, Sun J, Ng K, Yan X. (2019) Recurrent Neural Networks for Early Detection of Heart Failure From Longitudinal Electronic Health Record Data: Implications for Temporal Modeling With Respect to Time Before Diagnosis, Data Density, Data Quantity, and Data Type. Circ Cardiovasc Qual Outcomes, 12: e005114.
85. Rasmy L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F, Wu H, Xu H, Zhi D. (2018) A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. J Biomed Inform, 84: 11-16.
86. Nirschl JJ, Janowczyk A, Peyster EG, Frank R, Margulies KB, Feldman MD, Madabhushi A. (2018) A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue. PLOS ONE, 13: e0192726.

87. Liu Y, Guo X, Zheng Y. (2019) An Automatic Approach Using ELM Classifier for HFpEF Identification Based on Heart Sound Characteristics. Journal of Medical Systems, 43: 285.
88. Yoon S, Eom GH. (2019) Heart failure with preserved ejection fraction: present status and future directions. Experimental & Molecular Medicine, 51: 1-9.
89. Solomon SD, McMurray JJV, Anand IS, Ge J, Lam CSP, Maggioni AP, Martinez F, Packer M, Pfeffer MA, Pieske B, Redfield MM, Rouleau JL, van Veldhuisen DJ, Zannad F, Zile MR, Desai AS, Claggett B, Jhund PS, Boytsov SA, Comin-Colet J, Cleland J, Düngen H-D, Goncalvesova E, Katova T, Kerr Saraiva JF, Lelonek M, Merkely B, Senni M, Shah SJ, Zhou J, Rizkala AR, Gong J, Shi VC, Lefkowitz MP. (2019) Angiotensin–Neprilysin Inhibition in Heart Failure with Preserved Ejection Fraction. New England Journal of Medicine, 381: 1609-1620.
90. Pitt B, Pfeffer MA, Assmann SF, Boineau R, Anand IS, Claggett B, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Harty B, Heitner JF, Kenwood CT, Lewis EF, O'Meara E, Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, Yang S, McKinlay SM. (2014) Spironolactone for Heart Failure with Preserved Ejection Fraction. New England Journal of Medicine, 370: 1383-1392.
91. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, Brunner–La Rocca H-P, Choi D-J, Chopra V, Chuquiure-Valenzuela E, Giannetti N, Gomez-Mesa JE, Janssens S, Januzzi JL, Gonzalez-Juanatey JR, Merkely B, Nicholls SJ, Perrone SV, Piña IL, Ponikowski P, Senni M, Sim D, Spinar J, Squire I, Taddei S, Tsutsui H, Verma S, Vinereanu D, Zhang J, Carson P, Lam CSP, Marx N, Zeller C, Sattar N, Jamal W, Schnaidt S, Schnee JM, Brueckmann M, Pocock SJ, Zannad F, Packer M. (2021) Empagliflozin in Heart Failure with a Preserved Ejection Fraction. New England Journal of Medicine, doi:10.1056/NEJMoa2107038.
92. Shah AM, Pfeffer MA. (2012) The many faces of heart failure with preserved ejection fraction. Nat Rev Cardiol, 9: 555-556.
93. Shah AM, Solomon SD. (2012) Phenotypic and pathophysiological heterogeneity in heart failure with preserved ejection fraction. Eur Heart J, 33: 1716-1717.

94. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang CC, Deo RC. (2015) Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation, 131: 269-279.
95. Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, Yarde M, Wang Z, Bhattacharya PT, Chirinos DA, Prenner S, Zamani P, Seiffert DA, Car BD, Gordon DA, Margulies K, Cappola T, Chirinos JA. (2020) Clinical Phenogroups in Heart Failure With Preserved Ejection Fraction. JACC: Heart Failure, 8: 172-184.
96. Hedman Å K, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan LM, Shah SJ, Linde CM, Donal E, Daubert JC, Mälarstig A, Ziemek D, Lund L. (2020) Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. Heart, 106: 342-349.
97. Przewlocka-Kosmala M, Marwick TH, Dabrowski A, Kosmala W. (2019) Contribution of Cardiovascular Reserve to Prognostic Categories of Heart Failure With Preserved Ejection Fraction: A Classification Based on Machine Learning. Journal of the American Society of Echocardiography, 32: 604-615.e606.
98. Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL, Kitzman DW, Lee KL, O'Connor CM, Felker GM. (2014) Clinical implications of chronic heart failure phenotypes defined by cluster analysis. J Am Coll Cardiol, 64: 1765-1774.
99. Horiuchi Y, Tanimoto S, Latif AHMM, Urayama KY, Aoki J, Yahagi K, Okuno T, Sato Y, Tanaka T, Koseki K, Komiyama K, Nakajima H, Hara K, Tanabe K. (2018) Identifying novel phenotypes of acute heart failure using cluster analysis of clinical variables. International Journal of Cardiology, 262: 57-63.
100. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlström U, O'Connor CM, Felker GM, Desai NR. (2018) Machine Learning Methods Improve Prognostication, Identify Clinically Distinct Phenotypes, and Detect Heterogeneity in Response to Therapy in a Large Cohort of Heart Failure Patients. J Am Heart Assoc, 7.
101. Cho JS, Shrestha S, Kagiyama N, Hu L, Ghaffar YA, Casaclang-Verzosa G, Zeb I, Sengupta PP. (2020) A Network-Based "Phenomics" Approach for Discovering Patient Subtypes From High-Throughput Cardiac Imaging Data. JACC: Cardiovascular Imaging, 13: 1655-1670.

102. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, Hisamitsu T, Kojima G, Felsted J, Kakarmath S, Kvedar J, Jethwani K. (2018) A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak, 18: 44.
103. Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Dwivedi G. (2019) Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. ESC Heart Fail, 6: 428-435.
104. Liu W, Stansbury C, Singh K, Ryan AM, Sukul D, Mahmoudi E, Waljee A, Zhu J, Nallamothu BK. (2020) Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. PLOS ONE, 15: e0221606.
105. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, Negahban SN, Krumholz HM. (2016) Analysis of Machine Learning Techniques for Heart Failure Readmissions. Circ Cardiovasc Qual Outcomes, 9: 629-640.
106. Yu S, Farooq F, van Esbroeck A, Fung G, Anand V, Krishnapuram B. (2015) Predicting readmission risk with institution-specific prediction models. Artif Intell Med, 65: 89-96.
107. Samad MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, Williams BA, Haggerty CM, Fornwalt BK. (2019) Predicting Survival From Large Echocardiography and Electronic Health Record Datasets: Optimization With Machine Learning. JACC Cardiovasc Imaging, 12: 681-689.
108. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, Zhu W, Sama I, Tadel M, Campagnari C, Greenberg B, Yagil A. (2020) Improving risk prediction in heart failure using machine learning. European Journal of Heart Failure, 22: 139-147.
109. Hearn J, Ross HJ, Mueller B, Fan C-P, Crowdy E, Duhamel J, Walker M, Alba AC, Manlhiot C. (2018) Neural Networks for Prognostication of Patients With Heart Failure. Circulation: Heart Failure, 11: e005193.
110. Kwon JM, Kim KH, Jeon KH, Lee SE, Lee HY, Cho HJ, Choi JO, Jeon ES, Kim MS, Kim JJ, Hwang KK, Chae SC, Baek SH, Kang SM, Choi DJ, Yoo BS, Kim KH, Park HY, Cho MC, Oh BH. (2019) Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. PLoS One, 14: e0219302.

111. Taslimitehrani V, Dong G, Pereira NL, Panahiazar M, Pathak J. (2016) Developing EHR-driven heart failure risk prediction models using CPXR(Log) with the probabilistic loss function. J Biomed Inform, 60: 260-269.
112. Han JJ, Acker MA, Atluri P. (2018) Left Ventricular Assist Devices. Circulation, 138: 2841-2851.
113. Rogers JG, Pagani FD, Tatooles AJ, Bhat G, Slaughter MS, Birks EJ, Boyce SW, Najjar SS, Jeevanandam V, Anderson AS, Gregoric ID, Mallidi H, Leadley K, Aaronson KD, Frazier OH, Milano CA. (2017) Intrapericardial Left Ventricular Assist Device for Advanced Heart Failure. New England Journal of Medicine, 376: 451-460.
114. Mehra MR, Naka Y, Uriel N, Goldstein DJ, Cleveland JC, Colombo PC, Walsh MN, Milano CA, Patel CB, Jorde UP, Pagani FD, Aaronson KD, Dean DA, McCants K, Itoh A, Ewald GA, Horstmanshof D, Long JW, Salerno C. (2016) A Fully Magnetically Levitated Circulatory Pump for Advanced Heart Failure. New England Journal of Medicine, 376: 440-450.
115. Kanwar MK, Lohmueller LC, Kormos RL, Teuteberg JJ, Rogers JG, Lindenfeld J, Bailey SH, McIlvennan CK, Benza R, Murali S, Antaki J. (2018) A Bayesian Model to Predict Survival After Left Ventricular Assist Device Implantation. JACC Heart Fail, 6: 771-779.
116. Cowger J, Sundareswaran K, Rogers JG, Park SJ, Pagani FD, Bhat G, Jaski B, Farrar DJ, Slaughter MS. (2013) Predicting survival in patients receiving continuous flow left ventricular assist devices: the HeartMate II risk score. J Am Coll Cardiol, 61: 313-321.
117. Loghmanpour NA, Kormos RL, Kanwar MK, Teuteberg JJ, Murali S, Antaki JF. (2016) A Bayesian Model to Predict Right Ventricular Failure Following Left Ventricular Assist Device Therapy. JACC Heart Fail, 4: 711-721.
118. Matthews JC, Koelling TM, Pagani FD, Aaronson KD. (2008) The right ventricular failure risk score a pre-operative tool for assessing the risk of right ventricular failure in left ventricular assist device candidates. J Am Coll Cardiol, 51: 2163-2172.
119. Drakos SG, Janicki L, Horne BD, Kfoury AG, Reid BB, Clayson S, Horton K, Haddad F, Li DY, Renlund DG, Fisher PW. (2010) Risk factors predictive of right

ventricular failure after left ventricular assist device implantation. Am J Cardiol, 105: 1030-1035.

120. Goldenberg I, Kutyifa V, Klein HU, Cannom DS, Brown MW, Dan A, Daubert JP, Estes NAM, Foster E, Greenberg H, Kautzner J, Klempfner R, Kuniss M, Merkely B, Pfeffer MA, Quesada A, Viskin S, McNitt S, Polonsky B, Ghanem A, Solomon SD, Wilber D, Zareba W, Moss AJ. (2014) Survival with Cardiac-Resynchronization Therapy in Mild Heart Failure. New England Journal of Medicine, 370: 1694-1701.

121. Normand C, Kaye DM, Povsic TJ, Dickstein K. (2019) Beyond pharmacological treatment: an insight into therapies that target specific aspects of heart failure pathophysiology. Lancet, 393: 1045-1055.

122. Steffel J, Ruschitzka F. (2014) Superresponse to cardiac resynchronization therapy. Circulation, 130: 87-90.

123. Kutyifa V, Geller L, Bogyi P, Zima E, Aktas MK, Ozcan EE, Becker D, Nagy VK, Kosztin A, Szilagyi S, Merkely B. (2014) Effect of cardiac resynchronization therapy with implantable cardioverter defibrillator versus cardiac resynchronization therapy with pacemaker on mortality in heart failure patients: results of a high-volume, single-centre experience. Eur J Heart Fail, 16: 1323-1330.

124. Nagy KV, Széplaki G, Perge P, Boros AM, Kosztin A, Apor A, Molnár L, Szilágyi S, Tahin T, Zima E, Kutyifa V, Gellér L, Merkely B. (2018) Quality of life measured with EuroQol-five dimensions questionnaire predicts long-term mortality, response, and reverse remodelling in cardiac resynchronization therapy patients. Europace, 20: 1506-1512.

125. Canepa M, Fonseca C, Chioncel O, Laroche C, Crespo-Leiro MG, Coats AJS, Mebazaa A, Piepoli MF, Tavazzi L, Maggioni AP. (2018) Performance of Prognostic Risk Scores in Chronic Heart Failure Patients Enrolled in the European Society of Cardiology Heart Failure Long-Term Registry. JACC Heart Fail, 6: 452-462.

126. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, Pouleur AC, Knappe D, Biering-Sørensen T, Kutyifa V, Moss A, Stein K, Solomon SD, Bijnens B. (2019) Machine learning-based phenogrouping in heart

failure to identify responders to cardiac resynchronization therapy. Eur J Heart Fail, 21: 74-85.

127. Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, Field ME, Eckhardt LL, Page CD. (2018) Machine Learning Algorithm Predicts Cardiac Resynchronization Therapy Outcomes: Lessons From the COMPANION Trial. Circ Arrhythm Electrophysiol, 11: e005499.

128. Howell SJ, Stivland T, Stein K, Ellenbogen KA, Tereshchenko LG. (2021) Using Machine-Learning for Prediction of the Response to Cardiac Resynchronization Therapy. JACC: Clinical Electrophysiology, doi:doi:10.1016/j.jacep.2021.06.009.

129. Feeny AK, Rickard J, Patel D, Toro S, Trulock KM, Park CJ, LaBarbera MA, Varma N, Niebauer MJ, Sinha S, Gorodeski EZ, Grimm RA, Ji X, Barnard J, Madabhushi A, Spragg DD, Chung MK. (2019) Machine Learning Prediction of Response to Cardiac Resynchronization Therapy. Circulation: Arrhythmia and Electrophysiology, 12: e007316.

130. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJV, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WHW, Tsai EJ, Wilkoff BL. (2013) 2013 ACCF/AHA Guideline for the Management of Heart Failure. Circulation, 128: e240-e327.

131. Feeny AK, Rickard J, Trulock KM, Patel D, Toro S, Moennich LA, Varma N, Niebauer MJ, Gorodeski EZ, Grimm RA, Barnard J, Madabhushi A, Chung MK. (2020) Machine Learning of 12-Lead QRS Waveforms to Identify Cardiac Resynchronization Therapy Patients With Differential Outcomes. Circulation: Arrhythmia and Electrophysiology, 13: e008210.

132. Cai C, Tafti AP, Ngufor C, Zhang P, Xiao P, Dai M, Liu H, Noseworthy P, Chen M, Friedman PA, Cha YM. (2021) Using ensemble of ensemble machine learning methods to predict outcomes of cardiac resynchronization. J Cardiovasc Electrophysiol, 32: 2504-2514.

133. Galli E, Le Rolle V, Smiseth OA, Duchenne J, Aalen JM, Larsen CK, Sade EA, Hubert A, Anilkumar S, Penicka M, Linde C, Leclercq C, Hernandez A, Voigt JU, Donal E. (2021) Importance of Systematic Right Ventricular Assessment in

Cardiac Resynchronization Therapy Candidates: A Machine Learning Approach. J Am Soc Echocardiogr, 34: 494-502.

134. Hu S-Y, Santus E, Forsyth AW, Malhotra D, Haimson J, Chatterjee NA, Kramer DB, Barzilay R, Tulsky JA, Lindvall C. (2019) Can machine learning improve patient selection for cardiac resynchronization therapy? PLOS ONE, 14: e0222397.

135. Jing L, Cerna AEU, Good CW, Sauers NM, Schneider G, Hartzel DN, Leader JB, Kirchner HL, Hu Y, Riviello DM, Stough JV, Gazes S, Haggerty A, Raghunath S, Carry BJ, Haggerty CM, Fornwalt BK. (2020) A Machine Learning Approach to Management of Heart Failure Populations. JACC: Heart Failure, 8: 578-587.

136. Howard JP, Fisher L, Shun-Shin MJ, Keene D, Arnold AD, Ahmad Y, Cook CM, Moon JC, Manisty CH, Whinnett ZI, Cole GD, Rueckert D, Francis DP. (2019) Cardiac Rhythm Device Identification Using Neural Networks. JACC: Clinical Electrophysiology, 5: 576-586.

137. Weinreich M, Chudow JJ, Weinreich B, Krumerman T, Nag T, Rahgozar K, Shulman E, Fisher J, Ferrick KJ. (2019) Development of an Artificially Intelligent Mobile Phone Application to Identify Cardiac Devices on Chest Radiography. JACC Clin Electrophysiol, 5: 1094-1095.

138. Chudow JJ, Jones D, Weinreich M, Zaremski L, Lee S, Weinreich B, Krumerman A, Fisher JD, Ferrick KJ. (2021) A Head-to Head Comparison of Machine Learning Algorithms for Identification of Implanted Cardiac Devices. Am J Cardiol, 144: 77-82.

139. Krittanawong C, Rogers AJ, Johnson KW, Wang Z, Turakhia MP, Halperin JL, Narayan SM. (2021) Integration of novel monitoring devices with machine learning technology for scalable cardiovascular management. Nature Reviews Cardiology, 18: 75-91.

140. Bachtiger P, Plymen CM, Pabari PA, Howard JP, Whinnett ZI, Opoku F, Janering S, Faisal AA, Francis DP, Peters NS. (2020) Artificial Intelligence, Data Sensors and Interconnectivity: Future Opportunities for Heart Failure. Card Fail Rev, 6: e11.

141. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, Lancellotti P, Muraru

D, Picard MH, Rietzschel ER, Rudski L, Spencer KT, Tsang W, Voigt JU. (2015) Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr, 28: 1-39.e14.

142. Abraham WT, Fisher WG, Smith AL, Delurgio DB, Leon AR, Loh E, Kocovic DZ, Packer M, Clavell AL, Hayes DL, Ellestad M, Trupp RJ, Underwood J, Pickering F, Truex C, McAtee P, Messenger J. (2002) Cardiac resynchronization in chronic heart failure. N Engl J Med, 346: 1845-1853.
143. Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T, Carson P, DiCarlo L, DeMets D, White BG, DeVries DW, Feldman AM. (2004) Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med, 350: 2140-2150.
144. Lam CSP, Arnott C, Beale AL, Chandramouli C, Hilfiker-Kleiner D, Kaye DM, Ky B, Santema BT, Sliwa K, Voors AA. (2019) Sex differences in heart failure. European Heart Journal, 40: 3859-3868c.
145. Dewan P, Rorth R, Jhund PS, Shen L, Raparelli V, Petrie MC, Abraham WT, Desai AS, Dickstein K, Kober L, Mogensen UM, Packer M, Rouleau JL, Solomon SD, Swedberg K, Zile MR, McMurray JJV. (2019) Differential Impact of Heart Failure With Reduced Ejection Fraction on Men and Women. J Am Coll Cardiol, 73: 29-40.
146. Sridhar AR, Yarlagadda V, Parasa S, Reddy YM, Patel D, Lakkireddy D, Wilkoff BL, Dawn B. (2016) Cardiac Resynchronization Therapy: US Trends and Disparities in Utilization and Outcomes. Circ Arrhythm Electrophysiol, 9: e003108.
147. Lund LH, Braunschweig F, Benson L, Stahlberg M, Dahlstrom U, Linde C. (2017) Association between demographic, organizational, clinical, and socio-economic characteristics and underutilization of cardiac resynchronization therapy: results from the Swedish Heart Failure Registry. Eur J Heart Fail, 19: 1270-1279.
148. Chatterjee NA, Borgquist R, Chang Y, Lewey J, Jackson VA, Singh JP, Metlay JP, Lindvall C. (2017) Increasing sex differences in the use of cardiac

resynchronization therapy with or without implantable cardioverter-defibrillator. Eur Heart J, 38: 1485-1494.

149. Arshad A, Moss AJ, Foster E, Padeletti L, Barsheshet A, Goldenberg I, Greenberg H, Hall WJ, McNitt S, Zareba W, Solomon S, Steinberg JS. (2011) Cardiac resynchronization therapy is more effective in women than in men: the MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy) trial. J Am Coll Cardiol, 57: 813-820.

150. Zusterzeel R, Spatz ES, Curtis JP, Sanders WE, Selzman KA, Pina IL, Bao H, Ponirakis A, Varosy PD, Masoudi FA, Canos DA, Strauss DG. (2015) Cardiac resynchronization therapy in women versus men: observational comparative effectiveness study from the National Cardiovascular Data Registry. Circ Cardiovasc Qual Outcomes, 8: S4-11.

151. Varma N, Mittal S, Prillinger JB, Snell J, Dalal N, Piccini JP. (2017) Survival in Women Versus Men Following Implantation of Pacemakers, Defibrillators, and Cardiac Resynchronization Therapy Devices in a Large, Nationwide Cohort. J Am Heart Assoc, 6.

152. Yin FH, Fan CL, Guo YY, Zhu H, Wang ZL. (2017) The impact of gender difference on clinical and echocardiographic outcomes in patients with heart failure after cardiac resynchronization therapy: A systematic review and meta-analysis. PLoS One, 12: e0176248.

153. Hsich EM. (2019) Sex Differences in Advanced Heart Failure Therapies. Circulation, 139: 1080-1093.

154. Tokodi M, Shrestha S, Bianco C, Kagiyama N, Casaclang-Verzosa G, Narula J, Sengupta PP. (2020) Interpatient Similarities in Cardiac Function: A Platform for Personalized Cardiovascular Medicine. JACC Cardiovasc Imaging, 13: 1119-1132.

155. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF, 3rd, Dokainish H, Edvardsen T, Flachskampf FA, Gillebert TC, Klein AL, Lancellotti P, Marino P, Oh JK, Popescu BA, Waggoner AD. (2016) Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr, 29: 277-314.

156. Nicolau M, Levine AJ, Carlsson G. (2011) Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc Natl Acad Sci U S A, 108: 7265-7270.
157. Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, Bottinger EP, Dudley JT. (2015) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med, 7: 311ra174.
158. Nielson JL, Paquette J, Liu AW, Guandique CF, Tovar CA, Inoue T, Irvine KA, Gensel JC, Kloke J, Petrossian TC, Lum PY, Carlsson GE, Manley GT, Young W, Beattie MS, Bresnahan JC, Ferguson AR. (2015) Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. Nat Commun, 6: 8581.
159. Hinks T, Zhou X, Staples K, Dimitrov B, Manta A, Petrossian T, Lum P, Smith C, Ward J, Howarth P, Walls A, Gadola SD, Djukanović R. (2015) Multidimensional endotypes of asthma: topological data analysis of cross-sectional clinical, pathological, and immunological data. Lancet, 385 Suppl 1: S42.
160. Torres BY, Oliveira JH, Thomas Tate A, Rath P, Cumnock K, Schneider DS. (2016) Tracking Resilience to Infections by Mapping Disease Space. PLoS Biol, 14: e1002436.
161. Hwang D, Kim HJ, Lee SP, Lim S, Koo BK, Kim YJ, Kook W, Andreini D, Al-Mallah MH, Budoff MJ, Cademartiri F, Chinnaiyan K, Choi JH, Conte E, Marques H, de Araújo Gonçalves P, Gottlieb I, Hadamitzky M, Leipsic JA, Maffei E, Pontone G, Raff GL, Shin S, Lee BK, Chun EJ, Sung JM, Lee SE, Berman DS, Lin FY, Virmani R, Samady H, Stone PH, Narula J, Bax JJ, Shaw LJ, Min JK, Chang HJ. (2021) Topological Data Analysis of Coronary Plaques Demonstrates the Natural History of Coronary Atherosclerosis. JACC Cardiovasc Imaging, doi:10.1016/j.jcmg.2020.11.009.
162. Dindin M, Umeda Y, Chazal F. Topological Data Analysis for Arrhythmia Detection Through Modular Neural Networks. Springer International Publishing, Cham, 2020: 177-188.
163. Rich JD, Burns J, Freed BH, Maurer MS, Burkhoff D, Shah SJ. (2018) Meta-Analysis Global Group in Chronic (MAGGIC) Heart Failure Risk Score:

Validation of a Simple Tool for the Prediction of Morbidity and Mortality in Heart Failure With Preserved Ejection Fraction. J Am Heart Assoc, 7: e009594.

164. Simpson J, Jhund PS, Silva Cardoso J, Martinez F, Mosterd A, Ramires F, Rizkala AR, Senni M, Squire I, Gong J, Lefkowitz MP, Shi VC, Desai AS, Rouleau JL, Swedberg K, Zile MR, McMurray JJV, Packer M, Solomon SD. (2015) Comparing LCZ696 with enalapril according to baseline risk using the MAGGIC and EMPHASIS-HF risk scores: an analysis of mortality and morbidity in PARADIGM-HF. J Am Coll Cardiol, 66: 2059-2071.

165. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2017.

166. Tokodi M, Schwertner WR, Kovács A, Tősér Z, Staub L, Sárkány A, Lakatos BK, Behon A, Boros AM, Perge P, Kutyifa V, Széplaki G, Gellér L, Merkely B, Kosztin A. (2020) Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. Eur Heart J, 41: 1747-1756.

167. Gasparini M, Klersy C, Leclercq C, Lunati M, Landolina M, Auricchio A, Santini M, Boriani G, Proclemer A, Leyva F. (2015) Validation of a simple risk stratification tool for patients implanted with Cardiac Resynchronization Therapy: the VALID-CRT risk score. Eur J Heart Fail, 17: 717-724.

168. Höke U, Mertens B, Khidir MJH, Schalij MJ, Bax JJ, Delgado V, Ajmone Marsan N. (2017) Usefulness of the CRT-SCORE for Shared Decision Making in Cardiac Resynchronization Therapy in Patients With a Left Ventricular Ejection Fraction of ≤35. Am J Cardiol, 120: 2008-2016.

169. Khatib M, Tolosana JM, Trucco E, Borràs R, Castel A, Berruezo A, Doltra A, Sitges M, Arbelo E, Matas M, Brugada J, Mont L. (2014) EAARN score, a predictive score for mortality in patients receiving cardiac resynchronization therapy based on pre-implantation risk factors. Eur J Heart Fail, 16: 802-809.

170. Providencia R, Marijon E, Barra S, Reitan C, Breitenstein A, Defaye P, Papageorgiou N, Duehmke R, Winnik S, Ang R, Klug D, Gras D, Oezkartal T, Segal OR, Deharo JC, Leclercq C, Lambiase PD, Fauchier L, Bordachar P, Steffel J, Sadoul N, Piot O, Borgquist R, Agarwal S, Chow A, Boveda S. (2018)

Usefulness of a clinical risk score to predict the response to cardiac resynchronization therapy. Int J Cardiol, 260: 82-87.

171. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M. (2006) The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation, 113: 1424-1433.
172. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. (2011) Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res., 12: 2825-2830.
173. Tokodi M, Behon A, Merkel ED, Kovács A, Tősér Z, Sárkány A, Csákvári M, Lakatos BK, Schwertner WR, Kosztin A, Merkely B. (2021) Sex-Specific Patterns of Mortality Predictors Among Patients Undergoing Cardiac Resynchronization Therapy: A Machine Learning Approach. Front Cardiovasc Med, 8: 611055.
174. Parimbelli E, Marini S, Sacchi L, Bellazzi R. (2018) Patient similarity for precision medicine: A systematic review. J Biomed Inform, 83: 87-96.
175. Zhang P, Wang F, Hu J, Sorrentino R. (2014) Towards personalized medicine: leveraging patient similarity and drug similarity analytics. AMIA Jt Summits Transl Sci Proc, 2014: 132-136.
176. Sharafoddini A, Dubin JA, Lee J. (2017) Patient Similarity in Prediction Models Based on Health Data: A Scoping Review. JMIR Med Inform, 5: e7.
177. Ng K, Sun J, Hu J, Wang F. (2015) Personalized Predictive Modeling and Risk Factor Identification using Patient Similarity. AMIA Jt Summits Transl Sci Proc, 2015: 132-136.
178. Omar AM, Bansal M, Sengupta PP. (2016) Advances in Echocardiographic Imaging in Heart Failure With Reduced and Preserved Ejection Fraction. Circ Res, 119: 357-374.
179. Antman EM, Loscalzo J. (2016) Precision medicine in cardiology. Nat Rev Cardiol, 13: 591-602.
180. Mirnezami R, Nicholson J, Darzi A. (2012) Preparing for precision medicine. N Engl J Med, 366: 489-491.

181. Casaclang-Verzosa G, Shrestha S, Khalil MJ, Cho JS, Tokodi M, Balla S, Alkhouli M, Badhwar V, Narula J, Miller JD, Sengupta PP. (2019) Network Tomography for Understanding Phenotypic Presentations in Aortic Stenosis. JACC: Cardiovascular Imaging, 12: 236-248.

182. Sengupta PP, Shrestha S, Kagiyama N, Hamirani Y, Kulkarni H, Yanamala N, Bing R, Chin CWL, Pawade TA, Messika-Zeitoun D, Tastet L, Shen M, Newby DE, Clavel M-A, Pibarot P, Dweck MR, Larose É, Guzzetti E, Bernier M, Beaudoin J, Arsenault M, Côté N, Everett R, Jenkins WSA, Tribouilloy C, Dreyfus J, Mathieu T, Renard C, Gun M, Macron L, Sechrist JW, Lacomis JM, Nguyen V, Gay LG, Calabria HC, Ntalas I, Prendergast B, Rajani R, Evangelista A, Cavalcante JL. (2021) A Machine-Learning Framework to Identify Distinct Phenotypes of Aortic Stenosis Severity. JACC: Cardiovascular Imaging, doi:doi:10.1016/j.jcmg.2021.03.020.

183. De Keulenaer GW, Brutsaert DL. (2011) Systolic and diastolic heart failure are overlapping phenotypes within the heart failure spectrum. Circulation, 123: 1996-2005.

184. Pandey A, Kagiyama N, Yanamala N, Segar MW, Cho JS, Tokodi M, Sengupta PP. (2021) Deep-Learning Models for the Echocardiographic Assessment of Diastolic Dysfunction. JACC: Cardiovascular Imaging, doi:doi:10.1016/j.jcmg.2021.04.010.

185. Redfield MM, Chen HH, Borlaug BA, Semigran MJ, Lee KL, Lewis G, LeWinter MM, Rouleau JL, Bull DA, Mann DL, Deswal A, Stevenson LW, Givertz MM, Ofili EO, O'Connor CM, Felker GM, Goldsmith SR, Bart BA, McNulty SE, Ibarra JC, Lin G, Oh JK, Patel MR, Kim RJ, Tracy RP, Velazquez EJ, Anstrom KJ, Hernandez AF, Mascette AM, Braunwald E, RELAX Trial ft. (2013) Effect of Phosphodiesterase-5 Inhibition on Exercise Capacity and Clinical Status in Heart Failure With Preserved Ejection Fraction: A Randomized Clinical Trial. JAMA, 309: 1268-1277.

186. Redfield MM, Anstrom KJ, Levine JA, Koepp GA, Borlaug BA, Chen HH, LeWinter MM, Joseph SM, Shah SJ, Semigran MJ, Felker GM, Cole RT, Reeves GR, Tedford RJ, Tang WHW, McNulty SE, Velazquez EJ, Shah MR, Braunwald

E. (2015) Isosorbide Mononitrate in Heart Failure with Preserved Ejection Fraction. New England Journal of Medicine, 373: 2314-2324.

187. Raatikainen MJP, Arnar DO, Merkely B, Nielsen JC, Hindricks G, Heidbuchel H, Camm J. (2017) A Decade of Information on the Use of Cardiac Implantable Electronic Devices and Interventional Electrophysiological Procedures in the European Society of Cardiology Countries: 2017 Report from the European Heart Rhythm Association. Europace, 19: ii1-ii90.

188. Allen LA, Matlock DD, Shetterly SM, Xu S, Levy WC, Portalupi LB, McIlvennan CK, Gurwitz JH, Johnson ES, Smith DH, Magid DJ. (2017) Use of Risk Models to Predict Death in the Next Year Among Individual Ambulatory Patients With Heart Failure. JAMA Cardiol, 2: 435-441.

189. Sengupta PP, Kulkarni H, Narula J. (2018) Prediction of Abnormal Myocardial Relaxation From Signal Processed Surface ECG. J Am Coll Cardiol, 71: 1650-1660.

190. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. (2016) Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med, 44: 368-374.

191. Weiss JC, Natarajan S, Peissig PL, McCarty CA, Page D. (2012) Machine Learning for Personalized Medicine: Predicting Primary Myocardial Infarction from Electronic Health Records. AI Magazine, 33: 33-45.

192. Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, Doughty RN. (2013) Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J, 34: 1404-1413.

193. Ouwerkerk W, Voors AA, Zwinderman AH. (2014) Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. JACC Heart Fail, 2: 429-436.

194. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM. (2020) Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. JACC Heart Fail, 8: 12-21.

195. Banerjee M, Reynolds E, Andersson HB, Nallamothu BK. (2019) Tree-Based Analysis. Circ Cardiovasc Qual Outcomes, 12: e004879.
196. Hothorn T, Hornik K, Zeileis A. (2006) Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15: 651-674.
197. de Waard D, Manlucu J, Gillis AM, Sapp J, Bernick J, Doucette S, Tang A, Wells G, Parkash R. (2019) Cardiac Resynchronization in Women: A Substudy of the Resynchronization-Defibrillation for Ambulatory Heart Failure Trial. JACC Clin Electrophysiol, 5: 1036-1044.
198. Schuchert A, Muto C, Maounis T, Frank R, Ella RO, Polauck A, Padeletti L. (2013) Gender-related safety and efficacy of cardiac resynchronization therapy. Clin Cardiol, 36: 683-690.
199. Beela AS, Duchenne J, Petrescu A, Ünlü S, Penicka M, Aakhus S, Winter S, Aarones M, Stefanidis E, Fehske W, Willems R, Szulik M, Kukulski T, Faber L, Ciarka A, Neskovic AN, Stankovic I, Voigt JU. (2019) Sex-specific difference in outcome after cardiac resynchronization therapy. Eur Heart J Cardiovasc Imaging, 20: 504-511.
200. Linde C, Cleland JGF, Gold MR, Claude Daubert J, Tang ASL, Young JB, Sherfesee L, Abraham WT. (2018) The interaction of sex, height, and QRS duration on the effects of cardiac resynchronization therapy on morbidity and mortality: an individual-patient data meta-analysis. Eur J Heart Fail, 20: 780-791.
201. Varma N, Lappe J, He J, Niebauer M, Manne M, Tchou P. (2017) Sex-Specific Response to Cardiac Resynchronization Therapy: Effect of Left Ventricular Size and QRS Duration in Left Bundle Branch Block. JACC Clin Electrophysiol, 3: 844-853.
202. Tamargo J, Rosano G, Walther T, Duarte J, Niessner A, Kaski JC, Ceconi C, Drexel H, Kjeldsen K, Savarese G, Torp-Pedersen C, Atar D, Lewis BS, Agewall S. (2017) Gender differences in the effects of cardiovascular drugs. Eur Heart J Cardiovasc Pharmacother, 3: 163-182.
203. Zusterzeel R, Curtis JP, Caños DA, Sanders WE, Selzman KA, Piña IL, Spatz ES, Bao H, Ponirakis A, Varosy PD, Masoudi FA, Strauss DG. (2014) Sex-specific

mortality risk by QRS morphology and duration in patients receiving CRT: results from the NCDR. J Am Coll Cardiol, 64: 887-894.

204. Varma N, Manne M, Nguyen D, He J, Niebauer M, Tchou P. (2014) Probability and magnitude of response to cardiac resynchronization therapy according to QRS duration and gender in nonischemic cardiomyopathy and LBBB. Heart Rhythm, 11: 1139-1147.

205. Loring Z, Caños DA, Selzman K, Herz ND, Silverman H, MaCurdy TE, Worrall CM, Kelman J, Ritchey ME, Piña IL, Strauss DG. (2013) Left bundle branch block predicts better survival in women than men receiving cardiac resynchronization therapy: long-term follow-up of ~ 145,000 patients. JACC Heart Fail, 1: 237-244.

206. Biton Y, Zareba W, Goldenberg I, Klein H, McNitt S, Polonsky B, Moss AJ, Kutyifa V. (2015) Sex Differences in Long-Term Outcomes With Cardiac Resynchronization Therapy in Mild Heart Failure Patients With Left Bundle Branch Block. J Am Heart Assoc, 4.

207. Strauss DG, Selvester RH, Wagner GS. (2011) Defining left bundle branch block in the era of cardiac resynchronization therapy. Am J Cardiol, 107: 927-934.

208. Linde C, Ståhlberg M, Benson L, Braunschweig F, Edner M, Dahlström U, Alehagen U, Lund LH. (2015) Gender, underutilization of cardiac resynchronization therapy, and prognostic impact of QRS prolongation and left bundle branch block in heart failure. Europace, 17: 424-431.

209. Kloosterman M, Maass AH. (2020) Sex differences in optimal atrioventricular delay in patients receiving cardiac resynchronization therapy. Clin Res Cardiol, 109: 124-127.

210. Cheng A, Gold MR, Waggoner AD, Meyer TE, Seth M, Rapkin J, Stein KM, Ellenbogen KA. (2012) Potential mechanisms underlying the effect of gender on response to cardiac resynchronization therapy: insights from the SMART-AV multicenter trial. Heart Rhythm, 9: 736-741.

211. Cipriani M, Ammirati E, Landolina M, Oliva F, Ghio S, Rordorf R, Lunati M. (2015) Cumulative analysis on 4802 patients confirming that women benefit more than men from cardiac resynchronization therapy. Int J Cardiol, 182: 454-456.

212. Ousdigian KT, Borek PP, Koehler JL, Heywood JT, Ziegler PD, Wilkoff BL. (2014) The epidemic of inadequate biventricular pacing in patients with persistent or permanent atrial fibrillation and its association with mortality. Circ Arrhythm Electrophysiol, 7: 370-376.
213. Auricchio A, Gasparini M, Linde C, Dobreanu D, Cano Ó, Sterlinski M, Bogale N, Stellbrink C, Refaat MM, Blomström-Lundqvist C, Lober C, Dickstein K, Normand C. (2019) Sex-Related Procedural Aspects and Complications in CRT Survey II: A Multicenter European Experience in 11,088 Patients. JACC Clin Electrophysiol, 5: 1048-1058.
214. Hayes DL, Boehmer JP, Day JD, Gilliam FR, 3rd, Heidenreich PA, Seth M, Jones PW, Saxon LA. (2011) Cardiac resynchronization therapy and the relationship of percent biventricular pacing to symptoms and survival. Heart Rhythm, 8: 1469-1475.
215. Martínez-Sellés M, Doughty RN, Poppe K, Whalley GA, Earle N, Tribouilloy C, McMurray JJ, Swedberg K, Køber L, Berry C, Squire I. (2012) Gender and survival in patients with heart failure: interactions with diabetes and aetiology. Results from the MAGGIC individual patient meta-analysis. Eur J Heart Fail, 14: 473-479.
216. Stolfo D, Uijl A, Vedin O, Strömberg A, Faxén UL, Rosano GMC, Sinagra G, Dahlström U, Savarese G. (2019) Sex-Based Differences in Heart Failure Across the Ejection Fraction Spectrum: Phenotyping, and Prognostic and Therapeutic Implications. JACC Heart Fail, 7: 505-515.
217. Ghali JK, Krause-Steinrauf HJ, Adams KF, Khan SS, Rosenberg YD, Yancy CW, Young JB, Goldman S, Peberdy MA, Lindenfeld J. (2003) Gender differences in advanced heart failure: insights from the BEST study. J Am Coll Cardiol, 42: 2128-2134.
218. Chandramouli C, Teng TK, Tay WT, Yap J, MacDonald MR, Tromp J, Yan L, Siswanto B, Reyes EB, Ngarmukos T, Yu CM, Hung CL, Anand I, Richards AM, Ling LH, Regensteiner JG, Lam CSP. (2019) Impact of diabetes and sex in heart failure with reduced ejection fraction patients from the ASIAN-HF registry. Eur J Heart Fail, 21: 297-307.

10. BIBLIOGRAPHY OF THE CANDIDATE

10.1 Bibliography related to the present thesis

1. **Tokodi M**, Shrestha S, Bianco C, Kagiyama N, Casaclang-Verzosa G, Narula J, Sengupta PP
   Interpatient similarities in cardiac function: a platform for personalized cardiovascular medicine
   JACC Cardiovasc Imaging 2020;**13**(5): 1119-1132.
   DOI: 10.1016/j.jcmg.2019.12.018
   *IF: **14.805***

2. **Tokodi M***, Schwertner WR*, Kovács A, Tősér Z, Staub L, Sárkány A, Lakatos BK, Behon A, Boros AM, Perge P, Kutyifa V, Széplaki G, Geller L, Merkely B, Kosztin A
   Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score
   Eur Heart J 2020;**41**(18): 1747-1756.
   DOI: 10.1093/eurheartj/ehz902
   **Márton Tokodi, M.D. and Walter R. Schwertner, M.D. are joint first authors.*
   *IF: **29.983***

3. **Tokodi M***, Behon A*, Merkel ED, Kovács A, Tősér Z, Sárkány A, Csákvári M, Lakatos BK, Schwertner WR, Kosztin A, Merkely B
   Sex-specific patterns of mortality predictors among patients undergoing cardiac resynchronization therapy: a machine learning approach
   *Front Cardiovasc Med 2021;**8**: 611055.*
   DOI: 10.3389/fcvm.2021.611055
   **Márton Tokodi, M.D. and Anett Behon, M.D. are joint first authors.*
   *IF (2020): **6.050***

10.2 Bibliography not related to the present thesis

1. Lakatos BK, Tősér Z, **Tokodi M**, Doronina A, Kosztin A, Muraru D, Badano LP, Kovács A, Merkely B
   Quantification of the relative contribution of the different right ventricular wall motion components to right ventricular ejection fraction: the ReVISION method
   Cardiovasc Ultrasound 2017;**15**(1): 8. DOI: 10.1186/s12947-017-0100-0
   *IF: **1.652***

2. Lakatos BK, **Tokodi M**, Assabiny A, Tősér Z, Kosztin A, Doronina A, Rácz K, Koritsánszky KB, Berzsenyi V, Németh E, Sax B, Kovács A, Merkely B
   Dominance of free wall radial motion in global right ventricular function of heart transplant recipients
   Clin Transplant 2018;**32**(3): e13192. DOI: 10.1111/ctr.13192
   *IF: **1.667***

3. Mátyás C, Kovács A, Németh BT, Oláh A, Braun S, **Tokodi M**, Barta BA, Benke K, Ruppert M, Lakatos BK, Merkely B, Radovits T
   Comparison of speckle-tracking echocardiography with invasive hemodynamics for the detection of characteristic cardiac dysfunction in type-1 and type-2 diabetic rat models
   Cardiovasc Diabetol 2018;**17**(1): 13. DOI: 10.1186/s12933-017-0645-0
   *IF: **5.948***

4. Doronina A, Édes IF, Újvári A, Kántor Z, Lakatos BK, **Tokodi M**, Sydó N, Kiss O, Abramov A, Kovács A, Merkely B
   The female athlete's heart: comparison of cardiac changes induced by different types of exercise training using 3D echocardiography
   Biomed Res Int 2018;**2018**: 3561962. DOI: 10.1155/2018/3561962
   *IF: **2.197***

5. Lakatos BK, Kiss O, **Tokodi M**, Tősér Z, Sydó N, Merkely G, Babity M, Szilágyi M, Komócsin Z, Bognár C, Kovács A, Merkely B
   Exercise-induced shift in right ventricular contraction pattern: novel marker of athlete's heart?
   Am J Physiol Heart Circ Physiol 2018;**315**(6): H1640-H1648.
   DOI: 10.1152/ajpheart.00304.2018
   *IF: **4.048***

6. Oláh A, Kovács A, Lux A, **Tokodi M**, Braun S, Lakatos BK, Mátyás C, Kellermayer D, Ruppert M, Sayour AA, Barta BA, Merkely B, Radovits T
   Characterization of the dynamic changes in left ventricular morphology and function induced by exercise training and detraining
   Int J Cardiol 2019;**277**: 178-185. DOI: 10.1016/j.ijcard.2018.10.092
   *IF: **3.229***

7. Casaclang-Verzosa G, Shrestha S, Khalil MJ, Cho JS, **Tokodi M**, Balla S, Alkhouli M, Badhwar V, Narula J, Miller JD, Sengupta PP
   Network tomography for understanding phenotypic presentations in aortic stenosis
   JACC Cardiovasc Imaging 2019;**12**(2): 236-248.
   DOI: 10.1016/j.jcmg.2018.11.025
   *IF: **12.740***

8. Kovács A, Lakatos B, **Tokodi M**, Merkely B
   Right ventricular mechanical pattern in health and disease: beyond longitudinal shortening
   *Heart Fail Rev 2019;**24**(4): 511-520.* DOI: 10.1007/s10741-019-09778-1
   *IF: **3.538***

9. Lakatos BK, Molnár AÁ, Kiss O, Sydó N, **Tokodi M**, Solymossi B, Fábián A, Dohy Z, Vágó H, Babity M, Bognár C, Kovács A, Merkely B
   Relationship between cardiac remodeling and exercise capacity in elite athletes: incremental value of left atrial morphology and function assessed by three-dimensional echocardiography
   J Am Soc Echocardiogr 2020;**33**(1): 101-109.e1.
   DOI: 10.1016/j.echo.2019.07.017
   *IF: **5.251***

10. **Tokodi M**, Németh E, Lakatos BK, Kispál E, Tősér Z, Staub L, Rácz K, Soltész Á, Szigeti S, Varga T, Gál J, Merkely B, Kovács A
    Right ventricular mechanical pattern in patients undergoing mitral valve surgery: a predictor of post-operative dysfunction?
    ESC Heart Failure 2020;**7**(3): 1246-1256. DOI: 10.1002/ehf2.12682
    *IF: **4.411***

11. Lakatos BK, Nabeshima Y, **Tokodi M**, Nagata Y, Tősér Z, Otani K, Kitano T, Fábián A, Újvári A, Boros AM, Merkely B, Kovács A, Takeuchi M
    Importance of non-longitudinal motion components in right ventricular function: 3D echocardiographic study in healthy volunteers
    J Am Soc Echocardiogr 2020;**33**(8): 995-1005.e1.
    DOI: 10.1016/j.echo.2020.04.002
    *IF: **5.251***

12. Ruppert M, Lakatos BK, Braun S, **Tokodi M**, Karime C, Oláh A, Sayour AA, Hizoh I, Barta BA, Merkely B, Kovács A, Radovits T
    Longitudinal strain reflects ventriculo-arterial coupling rather than mere contractility in rat models of hemodynamic overload-induced heart failure
    J Am Soc Echocardiogr 2020;**33**(10): 1264-1275.e4.
    DOI: 10.1016/j.echo.2020.05.017
    *IF: **5.251***

13. Ujvári A, Lakatos BK, **Tokodi M**, Fábián A, Merkely B, Kovács A
    Evaluation of left ventricular structure and function using 3D echocardiography
    J Vis Exp 2020(164). DOI: 10.3791/61212
    *IF: **1.355***

14. Lakatos BK, **Tokodi M**, Kispál E, Merkely B, Kovács A
    Morphological and functional assessment of the right ventricle using 3D echocardiography
    J Vis Exp 2020(164). DOI: 10.3791/61214
    *IF : **1.355***

15. Fábián A, Lakatos BK, **Tokodi M**, Kiss AR, Sydó N, Csulak E, Kispál E, Babity M, Szűcs A, Kiss O, Merkely B, Kovács A
    Geometrical remodeling of the mitral and tricuspid annuli in response to exercise training: a 3-D echocardiographic study in elite athletes
    *Am J Physiol Heart Circ Physiol 2021;**320**(5): H1774-H1785.*
    DOI: 10.1152/ajpheart.00877.2020
    *IF (2020): **4.733***

16. Lakatos BK, Ruppert M, **Tokodi M**, Oláh A, Braun S, Karime C, Ladányi Z, Sayour AA, Barta BA, Merkely B, Radovits T, Kovács A
    Myocardial work index: a marker of left ventricular contractility in pressure- or volume overload-induced heart failure
    *ESC Heart Fail 2021;**8**(3): 2220-2231.* DOI: 10.1002/ehf2.13314
    *IF (2020): **4.411***

17. **Tokodi M**, Levente S, Budai Á, Lakatos BK, Csákvári M, Suhai FI, Szabó L, Fábián A, Vágó H, Tősér Z, Merkely B, Kovács A
    Partitioning the right ventricle into 15 segments and decomposing its motion using 3D echocardiography-based models: the updated ReVISION method
    *Front Cardiovasc Med 2021;**8**: 622118.* DOI: 10.3389/fcvm.2021.622118
    *IF (2020): **6.050***

18. **Tokodi M**, Lakatos BK, Ruppert M, Fábián A, Oláh A, Sayour AA, Ladányi Z, Soós A, Merkely B, Sengupta PP, Radovits T, Kovács A
    Left ventricular pressure-strain-volume loops for the noninvasive assessment of volume overload-induced myocardial dysfunction
    *JACC Cardiovasc Imaging 2021;**14**(9):1868-1871.*
    DOI: 10.1016/j.jcmg.2021.03.005
    *IF (2020): **14.805***

19. Pandey A, Kagiyama N, Yanamala N, Segar MW, Cho JS, **Tokodi M**, Sengupta PP
    Deep-learning models for the echocardiographic assessment of diastolic dysfunction
    *JACC Cardiovasc Imaging 2021;**14**(10):1887-1900.*
    DOI: 10.1016/j.jcmg.2021.04.010
    *IF (2020): **14.805***

20. Schwertner WR, Behon A, Merkel ED, **Tokodi M**, Kovács A, Zima E, Osztheimer I, Molnár L, Király Á, Papp R, Gellér L, Kuthi L, Veres B, Kosztin A, Merkely B
    Long-term survival following upgrade compared with de novo cardiac resynchronization therapy implantation: a single-centre, high-volume experience
    *Europace 2021;**23**(8):1310-1318.* DOI: 10.1093/europace/euab059
    *IF (2020): **5.214***

21. **Tokodi M**, Oláh A, Fábián A, Lakatos BK, Hizoh I, Ruppert M, Sayour AA, Barta BA, Kiss O, Sydó N, Csulak E, Ladányi Z, Merkely B, Kovács A, Radovits T
Novel insights into the athlete's heart: is myocardial work the new champion of systolic function?
Eur Heart J Cardiovasc Imaging 2021. DOI: 10.1093/ehjci/jeab162
IF (2020): ***6.875***

22. Molnár AÁ, Kolossváry M, Lakatos BK, **Tokodi M**, Tárnoki ÁD, Tárnoki DL, Kovács A, Szilveszter B, Vörös S, Jermendy G, Maurovich-Horvat P, Merkely B
Left ventricular systolic function has strong independent genetic background from diastolic function: a classical twin study
*Medicina 2021;**57**(9): 935.* DOI: 10.3390/medicina57090935
IF (2020): ***2.430***

23. Csulak E, Petrov Á, Kováts T, **Tokodi M**, Lakatos BK, Kovács A, Staub L, Suhai FI, Szabó EL, Dohy Z, Vágó H, Becker D, Müller V, Sydó N, Merkely B
The impact of COVID-19 on the preparation for the Tokyo Olympics: a comprehensive performance assessment of top swimmers
*Int. J. Environ. Res. Public Health 2021;**18**(18): 9770.*
DOI: 10.3390/ijerph18189770
IF (2020): ***3.390***

24. Surkova E, Kovács A, **Tokodi M**, Lakatos BK, Merkely B, Muraru D, Ruocco A, Parati G, Badano LP
Contraction patterns of the right ventricle associated with different degrees of left ventricular systolic dysfunction
Circ Cardiovasc Imaging 2021. DOI: 10.1161/CIRCIMAGING.121.012774
IF (2020): ***7.792***

Hungarian articles:

1. Lakatos BK, Kovács A, **Tokodi M**, Doronina A, Merkely B
   [Assessment of the right ventricular anatomy and function by advanced echocardiography: pathological and physiological insights]
   Orv Hetil 2016;**157**(29): 1139-46. DOI: 10.1556/650.2016.30491
   *IF: **0.349***

2. Ujvári A, Komka Z, Kántor Z, Lakatos BK, **Tokodi M**, Doronina A, Babity M, Bognár C, Kiss O, Merkely B, Kovács A
   [3D echocardiographic analysis of elite kayak/canoe athletes]
   Cardiol Hung 2018;**48**(1): 13-19.
   DOI: 10.26430/CHUNGARICA.2018.48.1.13

3. Fábián A, Lakatos BK, Kiss O, Sydó N, Vágó H, Czimbalmos C, **Tokodi M**, Kántor Z, Bognár C, Major D, Kovács A, Merkely B
   [Functional shift of right ventricular mechanics in athletes: a three-dimensional echocardiography study]
   Cardiol Hung 2019;**49**(1): 17-23.
   DOI: 10.26430/CHUNGARICA.2019.49.1.17

4. Parázs N, Lakatos BK, Kovács A, Assabiny A, Király Á, Tarjányi Z, Szakál-Tóth Z, Teszák T, **Tokodi M**, Ujvári A, Kugler S, Szücs N, Merkely B, Sax B
   [Right heart failure many years after heart transplantation – a case of a rare etiology]
   Cardiol Hung 2021;**51**(1): 69-72.
   DOI: 10.26430/CHUNGARICA.2021.51.1.69

11. ACKNOWLEDGEMENTS

Carrying out the requisite work and writing this Ph.D. thesis was undoubtedly the most arduous task I have ever undertaken. It has been a truly life-changing experience for me, and it would not have been possible without the tremendous support and guidance that I received from many people.

First and foremost, I am extremely grateful to my supervisor, Dr. Attila Kovács, for his invaluable advice, continuous support, and patience during my Ph.D. fellowship. Thank you for seeing my potential as a researcher and introducing me to the realm of echocardiography and research as a medical student. Your immense knowledge and plentiful experience have encouraged me at all times in my academic research and daily life.

I would like to also express my sincere gratitude to Prof. Béla Merkely for giving me the opportunity to carry out my research projects, providing the background required for my research, and sharing my enthusiasm for machine learning.

I am deeply grateful to Prof. Partho Sengupta for inviting me to join the Innovation Center at the West Virginia University and showing me new horizons in cardiovascular research. I am particularly grateful to Sirish Shrestha, who has been supporting me not only as a colleague but also as a friend, and who made my fellowship and life in the US a wonderful experience.

I am also grateful to Dr. Annamária Kosztin, whose clinical and scientific expertise was invaluable for my research projects, especially for those related to cardiac resynchronization therapy.

I would like to extend my sincere thanks to Dr. Bálint Lakatos and Dr. Alexandra Fábián for their unwavering support and friendship and for motivating me to become a better researcher and, ultimately, a better person.

I would like to offer my special thanks to Zoltán Tősér and the entire Argus team for providing me technical guidance on how to amalgamate cardiovascular medicine and data science. I am also grateful to András Soós and Bálint Magyar for showing me the world of deep learning and for the wonderful gastronomic experiences.

Many thanks to all Ph.D. students and student researchers, in particular Dr. Walter Schwertner, Dr. Anett Behon, Zsuzsanna Ladányi, Máté Tolvaj, Flóra Kulcsár, Erika

Kispál and all co-authors of my papers for their scientific contribution. Without their help, none of my achievements would have been possible.

Above all, I would like to thank my wife, Petra, for being a constant source of love and support, for all the late nights and early mornings, and for keeping me sane over the past few months. But most of all, thank you for being my soulmate, the love of my life, and my better half.

I am highly grateful to my friends for being a source of inspiration that has always instilled confidence in me and has kept me motivated.

Last but not least, I would like to express my gratitude to my parents and family for being my backbone in life and providing limitless love, support, and care. Words can never express how much I am grateful to them.