Exploring Guidelines for Classification of Major Heart Failure Subtypes by Using Machine Learning

Amparo Alonso-Betanzos[1], Verónica Bolón-Canedo[1], Guy R. Heyndrickx[2] and Peter L.M. Kerkhof[3]

[1]Department of Computer Science, Universidad de A Coruña, Coruña, Spain. [2]Cardiovascular Center Aalst, OLV Clinic, Aalst, Belgium. [3]Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, The Netherlands.

Supplementary Issue: Heart Failure: An Exploration of Recent Advances in Research and Treatment

ABSTRACT

BACKGROUND: Heart failure (HF) manifests as at least two subtypes. The current paradigm distinguishes the two by using both the metric ejection fraction (EF) and a constraint for end-diastolic volume. About half of all HF patients exhibit preserved EF. In contrast, the classical type of HF shows a reduced EF. Common practice sets the cut-off point often at or near EF = 50%, thus defining a linear divider. However, a rationale for this *safe* choice is lacking, while the assumption regarding applicability of strict linearity has not been justified. Additionally, some studies opt for eliminating patients from consideration for HF if $40 < EF < 50\%$ (gray zone). Thus, there is a need for documented classification guidelines, solving gray zone ambiguity and formulating crisp delineation of transitions between phenotypes.

METHODS: Machine learning (ML) models are applied to classify HF subtypes within the ventricular volume domain, rather than by the single use of EF. Various ML models, both unsupervised and supervised, are employed to establish a foundation for classification. Data regarding 48 HF patients are employed as training set for subsequent classification of Monte Carlo–generated surrogate HF patients ($n = 403$). Next, we map consequences when EF cut-off differs from 50% (as proposed for women) and analyze HF candidates not covered by current rules.

RESULTS: The training set yields best results for the Support Vector Machine method (test error 4.06%), covers the gray zone, and other clinically relevant HF candidates. End-systolic volume (ESV) emerges as a logical discriminator rather than EF as in the prevailing paradigm.

CONCLUSIONS: Selected ML models offer promise for classifying HF patients (including the gray zone), when driven by ventricular volume data. ML analysis indicates that ESV has a role in the development of guidelines to parse HF subtypes. The documented curvilinear relationship between EF and ESV suggests that the assumption concerning a linear EF divider may not be of general utility over the complete clinically relevant range.

KEYWORDS: heart failure phenotype, ejection fraction, support vector machine, volume regulation graph

SUPPLEMENT: Heart Failure: An Exploration of Recent Advances in Research and Treatment

CITATION: Alonso-Betanzos et al. Exploring Guidelines for Classification of Major Heart Failure Subtypes by Using Machine Learning. *Clinical Medicine Insights: Cardiology* 2015:9(S1) 57–71 doi: 10.4137/CMC.S18746.

RECEIVED: November 30, 2014. **RESUBMITTED:** February 25, 2015. **ACCEPTED FOR PUBLICATION:** March 03, 2015.

ACADEMIC EDITOR: Thomas E. Vanhecke, Editor in Chief

TYPE: Perspective

FUNDING: This research has been economically supported in part by the Ministerio de Economía y Competitividad of the Spanish Government through the research project TIN 2012–37954, partially funded by FEDER funds of the European Union, and by the Consellería de Industria of the Xunta de Galicia through the research project GRC2014/035. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: ciamparo@udc.es, veronica.bolon@udc.es

Paper subject to independent expert blind peer review by minimum of two reviewers. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

Introduction

As George Pickering's reflections on the concept of hypertension show, the establishment of cut-off points can be seen as the result of a negotiation process involving all sorts of assumptions, both philosophical and purely practical.

[cited by Amsterdamska and Hiddinga][1]

Heart failure (HF) is a relatively common cardiac syndrome known for its severe sequelae, including death. The diagnosis is often only evident from the combination of symptoms (such as fatigue and dyspnea) and signs (eg, ankle edema), plus clinical investigations including the determination of left ventricular (LV) size and chamber filling pressure, in addition to information derived from specific biomarkers. In the United States, one in nine deaths in 2009 included HF as contributing cause. About half of people who develop HF die within 5 years of diagnosis. HF costs the nation an estimated $32 billion each year. This total includes the cost of health care services, medications to treat HF, and missed days of work.[2]

HF manifests as at least two subtypes, which are commonly distinguished on the basis of the metric ejection fraction (EF).[3–8] HF patients are typically designated into EF categories based on this single measurement at a point in time, although EF is not necessarily static.[5] Approximately half of all patients with HF have preserved ejection fraction (HFpEF), and they often include women and elderly.[9] Thus, as life expectancies continue to increase in western societies, the prevalence of HFpEF will continue to grow. Many features of the HF syndrome are similar across the EF spectrum,

including elevated left atrial pressure, abnormal LV-filling dynamics, neurohumoral activation, dyspnea, impaired exercise tolerance, frequent hospitalization, and reduced survival.[10] However, in contrast to (classical) HF with reduced ejection fraction (HFrEF), only a limited spectrum of treatment modalities seem effective in improving morbidity and mortality rates in HFpEF,[6] apart from measures for the reduction of risk factors (such as hypertension and obesity) plus a supportive approach (including diet, diuretics, and rehabilitation), and management of concomitant conditions (such as sleep apnea, insomnia, depression, and sexual dysfunction).[11] Additionally, a large pathophysiological heterogeneity exists within the broad spectrum of HFpEF.[12] Studies are further complicated by the existence of various comorbidities such as diabetes mellitus and kidney failure.

Traditionally, the cardiac performance index EF has been widely applied to conveniently assess the severity of cardiac problems.[13] In the particular case of the HF syndrome, it is clear that EF is only one of the many indicators to characterize the various aspects.[12] Typically, a low value of EF corresponds with serious cardiac problems and a poor prognosis. Calculation of EF is carried out by taking the ratio of two LV volume determinations during a cardiac cycle, namely, at the completion of filling and at maximal contraction. This ratio yields a dimensionless number, which obviously does not require (time-consuming) calibration of the underlying volumetric data. Although the calculation of EF is attractive from a practical point of view,[14] it is also evident from simple theoretical considerations that EF cannot be a valid indicator for severity in all types of heart diseases.[12,13] This notion is strikingly illustrated by recent observations showing that half of all HF patients do not exhibit the expected low value of EF. In contrast, such HF patients show higher values for EF, often comparable to those encountered in healthy individuals.[12] Advised cut-off levels to distinguish HFrEF from HFpEF are clearly formulated,[15,16] but unfortunately vary between 40% and 50% in actual clinical studies,[11] often making it difficult to interpret and compare their outcomes.[12] The reported frequent occurrence of transitions between major phenotypes of HF[5,17] also necessitates well-defined and uniform criteria for the separating line.[12]

The remarkable discrepancy concerning a numerical expression of EF and the two-fold routes for interpretation (ie, a normal value for EF may mean either no substantial cardiac disease or severe cardiac syndrome) urges further investigation. This dilemma cannot just be *solved* by the current practice of assigning a new name to a syndrome where the pertinent EF value seems to deny the presence of severe LV failure. Thus, *in terms of the traditional interpretation of EF*, we notice that the present terminology for the subdivision in a classical type (HFrEF) versus the new syndrome (indicated as HFpEF) is not sound or clarifying.[12] This dilemma can probably only be solved if we analyze LV volumes rather than their ratio (as reflected by EF).[13,14,18] We propose to leave the EF-centric paradigm and enter the Volume Regulation Graph (VRG) domain.[12] The latter includes iso-EF lines, which, however, are not coincident with regression lines for HFpEF and HFrEF.

Subgroups of HF patients are located in at least two distinct regions on the basis of their end-systolic volume (ESV) and EDV, and therefore uniquely located within the LV volume domain.[12,13] Our present study uses modern approaches to explore a more rational foundation for classifying two phenotypes of HF, in particular, by applying Machine Learning (ML) techniques.[19,20]

In an earlier study, we reported the remarkable connection between EF and one of its basic elements, namely, ESV, in contrast to the other constituent, ie, end-diastolic volume (EDV).[21] Following the definition,

$$EF = (1 - ESV/EDV) \times 100\%, \quad (1)$$

we notice that ESV and EDV form the building blocks for the analysis of LV mechanical function. Essentially, ESV and EDV are primary determinants of LV volume regulation.[12] Therefore, in the past, we had proposed to construct a graph of ESV versus EDV,[18] which has the clear advantage of yielding (nearly perfect) linear relationships.[12] Moreover, the index EF is implicitly incorporated in this representation. Indeed, iso-EF lines can be inscribed, leading to the notion that patient groups can theoretically be distinguished on the basis of their volume regulation characteristics.[12] Accordingly, the present study explores novel guidelines to classify patients as either HFpEF or HFrEF, while obviating the limitations inherent to EF by instead applying the paradigm of the VRG.[12] In summary, we address three relevant issues regarding classification of HF patients: consequences of varying the cut-off values, implications for borderline patients (in the gray zone), and proposals for categorizing an uncommon group of patients not covered by current guidelines and located in the region where EF >50% and end-diastolic volume index (EDVI) slightly greater than 97 mL/m^2.

Acceptance Rate of the Current Paradigm Concerning HF

In terms of LV volumetric characteristics, the two major HF phenotypes are distinguished on the basis of both EF and EDVI.[15,16] The present EF–EDVI model has several limitations, while the guidelines are not always adhered to:

1. There is no documentation why a cut-off for EF at 50% is the best choice.[10] In fact, all groups of investigators adopt their own cut-off at 40%, 45%, or 50% without clear motivation. It is often not explained why the official guidelines were NOT accepted.
2. Several researchers eliminate the discussion sub #1 by excluding patients with 40% < EF < 50%. The introduction of a gray zone is not proposed in guidelines,[11,15,16,22]

yet invoked by some authors without elaborating on the consequences of this particular choice.

3. A fair investigation concerning transition for HFrEF to HFpEF (and vice versa) requires a uniform statement regarding the dividing line. This especially applies to the above-mentioned gray zone where most likely the majority of the transitions occur.[5,10,17]
4. A small group of HF is not covered by the present guidelines, namely, those with EF >50% and EDVI >97, eg, with EDVI range extended to 102 mL/m^2.[7]

Thus, there is a need to refine or even reconsider the paradigm. This study explores new routes and opens up the discussion. ML results may show that a varying EF cut-off has advantages and that ESV may be preferred above EF, as suggested in Figure 1. We hope that these proposals will be evaluated in future clinical studies. A paradigm shift seems indicated because the dynamic metric EF depends on almost everything that affects heart function.[5,12–14,23–26]

Experimental Framework

Rather than looking at EF alone (as is done in most studies), we investigate HF from the perspective of the volume domain description.[12] This approach has the advantage of considering the basic variables that define the derived index EF, namely ESV and EDV.[18] An important impetus of the volume domain method resides in the fact that *stratification of patients* can easily be incorporated, as illustrated elsewhere.[12] From basic physiology, it is clear that ESV and EDV are highly linearly related for well-defined patient groups.[27] EF is inversely and nonlinearly connected with ESV,[12,13] while there is generally no significant relationship between EF and EDV.[21]

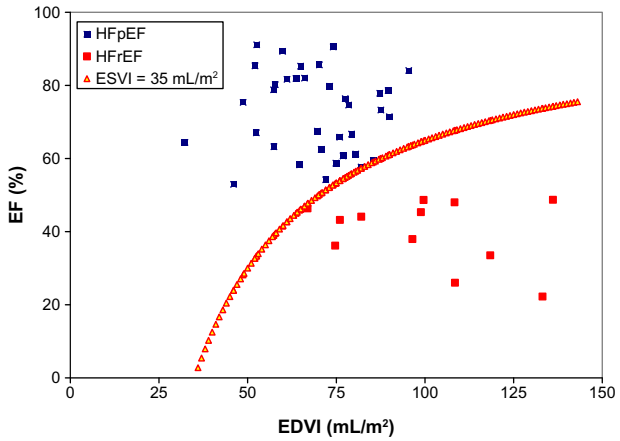

Figure 1. Scatter plot of all HF patients described in this study, relating EF to EDVI. The graph illustrates the connection between the old paradigm (relating EF to EDVI) and the new concept which highlights the exclusive importance of ESVI as indicated by the red-yellow curve. The European Society for Cardiology (ESC) guidelines currently require two constraints (ie, EF >50% and EDVI >97 mL/m^2). Our new approach indicates that one criterion may be sufficient, namely ESVI (here with cut-off set at 35 mL/m^2). As clearly illustrated in the figure, the old paradigm already includes the novel candidate. This graph immediately visualizes the alternative (and more simple) route to analyze and distinguish HF phenotypes, namely on the basis of ESVI.

The EF-related criterion in vogue today for HFrEF and the preserved EF variant of HF does not seem to be founded on a solid basis.[12,28–30] Thresholds adopted for EF vary from 40% to 50%.[4,11] Apart from the lack of consensus, we can discern additional shortcomings inherent to the present practice. A division between HFrEF and HFpEF based on a cut-off at 50% seems arbitrary. Clearly, patients with EFs of 51% and 49% are not different, and likely to cross back and forth the threshold,[10] resulting in frequent transitions.[5,17] A breakpoint value of 50% for EF simply appears to be selected halfway the spectrum, while an additional boundary condition for LV filling volume (ie, EDV <97 mL/m^2) in HFpEF may turn out to be superfluous. To resolve these issues we undertake an investigation including ML, specifically Support Vector Machine (SVM) approaches.[31] We analyzed actual HF patients plus an additional data set created by generating random numbers and subdividing them on the basis of theoretical guidelines. We have performed some experiments using the Weka tool,[32] which is a collection of ML algorithms and data preprocessing tools.[33] After that, the method that had a better generalization behavior, namely, SVM PEGASOS (Primal Estimated sub-GrAdient SOlver for SVM), was selected for carrying out the rest of the experiments.

In this retrospective HF study, we have included 35 patients with preserved EF (referred to as HFpEF) and 13 patients with reduced EF (denoted as HFrEF). LV volume and pressure data were obtained by one of the authors (GH) during diagnostic catheterization procedures at the Cardiovascular Center, OLV Clinic, Aalst, in Belgium, as described previously.[13] All 48 patients had an end-diastolic pressure above 16 mmHg, which is considered the gold standard, as opposed to tissue Doppler surrogates. Cardiac volumes are corrected for body surface area and indicated as such by the affix index (*I*). This normalization procedure is needed to ascertain compliance with the European Society for Cardiology (ESC) guidelines for classification of HF.[16] Further details regarding these patients and earlier clinical investigations regarding these groups have been reported elsewhere.[13]

One convenient route to evaluate various advanced approaches applies randomly generated numbers (*Monte Carlo* [MC] *simulation*). We created a database initially consisting of 63 data pairs and later expanded to 1000 data pairs, where the equivalent of EDVI was chosen to be the higher number and the end-systolic volume index (ESVI) the lower one in each pair. Subsequently, all data pairs beyond the (patho) physiological range were eliminated. Notably, data pairs were removed when calculated EF would refer to the hyperkinetic heart (EF >95%) or to near-terminal situations (EF <10%) or create excessive cardiac output. As a result, a working set of 403 data pairs remained, which reflect the full pathophysiological spectrum. These data pairs are explicitly assumed

to represent full-fledged surrogate HF patients not only in terms of their LV volume characteristics[13] but also supposed to comply with all clinically relevant requirements inherent to typical HF inclusion criteria, ie, elevated levels of specific biomarkers, and LV filling pressure above 16 mmHg. These assumptions provide the opportunity to fully focus on volume-related aspects of the HF syndrome and explore guidelines for classification.[12]

Results

Based on the Frank-Starling law, LV function is often described by relating stroke volume (SV) to EDV.[12,18] The ratio of SV and EDV yields EF (see Eq. 1). The combination of EF and EDV was traditionally proposed to specify criteria for HF subtypes[12] and may be referred to as the classical paradigm to distinguish HF patients. To illustrate this viewpoint, we present in Figure 1 the values of EF and EDVI for all our real patients. Just by eye-balling, it is evident that a fixed value of ESVI (see the yellow curve composed of the triangles) separates the two phenotypes of HF. This remarkable finding will be pursued in our analysis and the potential importance of ESVI as pivot will therefore be explored while applying ML methods.

Thus, the starting point of our present analysis is different and illustrated in Figure 2 by depicting the same real HF patients in a graph that relates the basic volumetric data, ie, ESVI and EDVI. The fundamental question now is how to find a solid method to separate the two phenotypes involved. In addition, consideration is given to the consequences of including a gray zone for EF values near the most frequently used cut-off levels (cf. gray triangle in Fig. 2 which encompasses the region where $40\% < EF < 50\%$).[12]

As a first step, we will analyze the patient data while utilizing the standard criteria (ie, cut-off at EF = 50% and EDV <97 mL/m^2 for HFpEF). Linear regression analysis for the two HF patient groups yields:

$$\text{HFrEF: ESVI} = 0.65\ \text{EDVI} - 5.44,\ r^2 = 0.894,\ \text{aveEF} = 39.84,\ n = 13 \quad (2)$$

$$\text{HFpEF: ESVI} = 0.31\ \text{EDVI} - 2.55,\ r^2 = 0.258,\ \text{aveEF} = 72.68,\ n = 35 \quad (3)$$

The slopes are significantly ($P < 0.005$) different, in contrast to EF versus EDVI for each group (see Fig. 1). Notice that n is the number of points and r^2 is the coefficient of determination, while aveEF is the average value of EF.

If data pairs within the gray zone (here for symmetry reasons selected around the EF = 50% line and thus limited to $45\% < EF < 55\%$) are eliminated, we find that linear regression analysis yields slightly different equations:

$$\text{HFrEF: ESVI} = 0.65\ \text{EDVI} + 0.47,\ r^2 = 0.939,\ \text{aveEF} = 35.13,\ n = 8 \quad (4)$$

$$\text{HFpEF: ESVI} = 0.34\ \text{EDVI} - 5.53,\ r^2 = 0.315,\ \text{aveEF} = 73.82,\ n = 33 \quad (5)$$

Essentially, the new set of regression lines results from the fact that the center zone around the iso-EF line at 50% has been selectively eliminated (cf. gray triangle in Fig. 2, referring to $40 < EF < 50\%$), as also demonstrated by the enhanced divergence of the new averages for EF in both groups.

The VRG-related regression coefficients for all applied EF cut-off levels are shown in Table 1.

An important observation is that the regression lines for both HF groups tend to run more in parallel as the cut-off value for EF is further reduced. However, at the same time, the values for the intercepts diverge, indicating that the LV

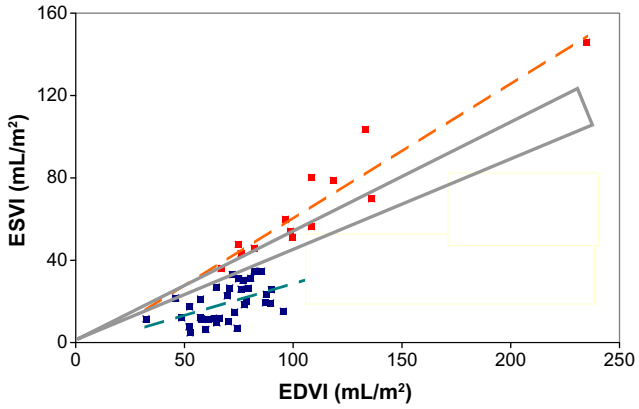

Figure 2. Scatter diagram for all 48 HF patients analyzed in this study. The superimposed gray triangle refers to the gray zone for EF, ie, the area where usually $40 < EF < 50\%$. Here the triangle is chosen slightly larger so as to include the higher range suggested for women. A graphical representation of ESVI versus EDVI of the left ventricle forms a logical combination, since major criteria for HF subtypes are based on volume-derived data. The regression lines for HFrEF and HFpEF are depicted in orange and green, respectively. Clearly, they refer to different and mutually exclusive areas in the LV volume domain. The gray triangle acts like a wedge to separate the two groups.

Table 1. Regression lines for different EF cut-offs (at 40%, 45%, 50%, and 55%).

EF (%)	ESVI–EDVI REGRESSION		R^2	N
	SLOPE	INTERCEPT		
55	HFrEF: 0.67	–8.00	0.91	15
	HFpEF: 0.34	–5.53	0.31	33
50	HFrEF: 0.65	–5.44	0.89	13
	HFpEF: 0.31	–2.55	0.26	35
45	HFrEF: 0.65	0.46	0.94	8
	HFpEF: 0.58	–19.78	0.58	40
40	HFrEF: 0.60	8.69	0.94	6
	HFpEF: 0.60	–19.91	0.55	42

volume regulation properties as reflected by the ESVI–EDVI diagram persistently remain different for both groups.

Results of applying ML techniques. ML is an area of Artificial Intelligence that explores algorithms and processes that are capable of learning from data.[34] Generally speaking, the data sets that should be available contain the values of a number of variables, which are called attributes. If there is a specially designated attribute, called output, and the aim of the process is to use a learning algorithm that predicts its value for instances that have not been seen yet, the process is called supervised learning, and a common task is classification, where the value to be predicted is a label. The set of instances used for the learning process is called training set, while the data set of unseen samples is called test set. If on the contrary, the process is aimed at extracting the information from the training data set, but without using the values of the output variable, the process is called unsupervised learning, and one of the most common methods is clustering, which aims at finding groups of items that are similar. There have been previous attempts at applying ML techniques to the prediction of HF[35] using variable selection methods, such as logistic regression and boosting, and to the prediction of HF subtypes,[36] although resulting prediction rates were relatively low. Other authors[37] explored whether clustering analysis using phenotypic data could identify phenotypically distinct HFpEF categories. However, our objective in this study is to use ML techniques to be able to discriminate between the patients with preserved EF and those with reduced EF using the concept of the VRG and exploring also some possible new guidelines for classification, opening up the discussion.

Our first experiments aimed at using unsupervised ML methods (ie, a method that does not make use of the knowledge regarding the output class to which the data points belong) for the three data sets provided, namely:

1. Data set 1: data from real patients, a total of 48 instances where 35 belong to class HFpEF and 13 to class HFrEF.
2. Data set 2: data simulated with MC, a total of 63 instances where 34 belong to class HFpEF and 29 to class HFrEF.
3. Data set 3: MC data generated as testing data, a total of 403 instances where 150 refer to class HFpEF, 137 belong to class HFrEF, while a third group (n = 116) still requires classification because on the basis of current guidelines they belong to neither HFpEF nor HFrEF. The third group is specifically introduced to challenge the universal validity of the current EF–EDVI paradigm (Fig. 1) which favors a linear separator based upon a fixed value for EF.

Unsupervised learning, as stated above, tries to find hidden structure in unlabeled data. Thus, as the examples feeding the algorithm are unlabeled, there is no error or reward signal to evaluate a potential solution. There are several algorithms that can be used for this purpose. Clustering is a common technique, consisting of grouping a set of data in such a way that those belonging to the same group (called a cluster) are more similar (in one sense or another, that is defined by the type of algorithm and its parameters) to each other than to those in other clusters. To perform an unsupervised separation of the two major phenotypes of HF patients, we evaluated three different clustering algorithms, using different approaches, all implemented in the Weka software tool[32]:

1. K-means using Euclidean distance, one of the most popular clustering methods,[38]
2. Expectation maximization (EM), which assigns a probability distribution to each instance indicating the probability of it belonging to each of the clusters,[39] and
3. Sequential information bottleneck algorithm (sIB), which assigns for each instance the cluster that has the minimum cost/distance to the instance.[40]

The results are shown in Figure 3 for the data sets 1 and 2 (real patients and simulated MC, including only the two major types of patient subgroups). Note that instances incorrectly assigned to a cluster are represented with a square in the figure and that only the third algorithm (sIB) tried to separate the samples using a similar approach as the current clinical guidelines.[16] However, it can also be seen that the patients reclassified in an alternative manner (see squares in Fig. 3) are all located within a region which in some other studies is conveniently neglected and benevolently referred to as gray zone.

Our second series of experiments employs a supervised automatic classification of both major HF types. Supervised learning infers a function from the analysis of labeled training data, and thus the real output for all data examples in the training set is available and used by the algorithm. The function inferred can subsequently be used for mapping new examples. An optimal scenario will allow the algorithm to correctly determine the class labels (ie, HF phenotype) for any new unseen instances. There are several classification algorithms that can be possibly used, each with its strengths and weaknesses, as no single algorithm obtains best results on all supervised learning problems. Thus, the performance of classification methods (ie, the adequacy of the resulting classification function) should be evaluated on a test set, different from the training set.

In order to decide which classifier is the best for the problem at hand, first several classifiers were tried, all belonging to different families, and using the actual patient data (48 instances) as training data with a 10-fold cross-validation, and the simulated MC data (n = 63) as testing data set. This division is a common practice in ML, as both data sets are relatively small, and it is recommended to use the real data for training, and obtaining adequate generalization.

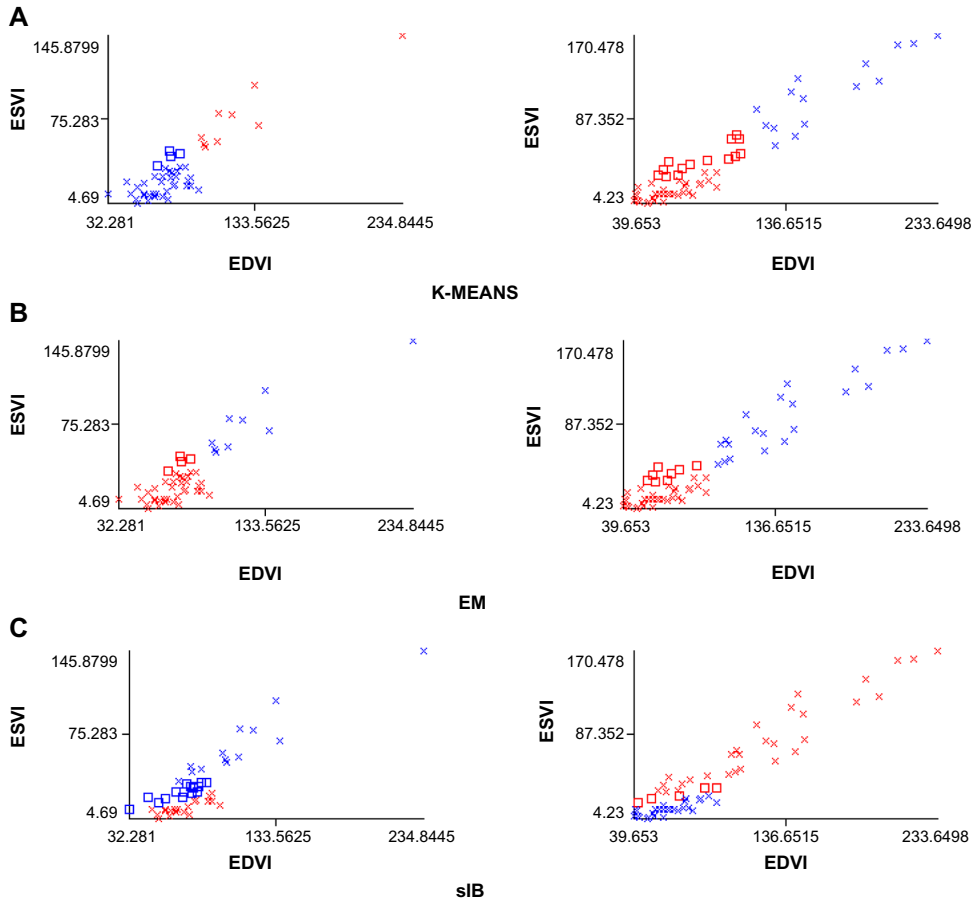

Figure 3. Results of three clustering algorithms [(**A**) K-means using Euclidean distance, (**B**) EM, and (**C**) sIB, respectively] regarding data set 1 (left) and data set 2 (right). The graphs display ESVI (on the y axis) versus EDVI (on the x axis). The axes are scaled to minimum and maximum values of the collection of data points. Squares represent instances that are incorrectly assigned to a cluster. Interestingly, the algorithms generate dividing patterns which deviate from the wedge paradigm shown in Figure 2.

The following classifiers, available at Weka tool, were tested:

1. Support Vector Machine methods: SVM PEGASOS[41] and SMO, which implements a Sequential Minimal Optimization algorithm for training a support vector classifier.[42]
2. Nearest neighbor classifiers (IB1),[43] which is a classic algorithm that uses normalized Euclidean distance to find the training instance closest to the given test instance, and predicts the same class as this training instance; and NNGE, a nearest neighbor-like algorithm using non-nested generalized exemplars.[44]
3. Rule-based algorithm OneR,[45] which is a simple, yet useful classification algorithm that generates one rule for each predictor in the data, and then selects the rule with the smallest total error as its *one rule*.
4. Tree-based algorithms (classic C4.5,[46] and PART,[47] that builds a partial C4.5 decision tree in each iteration and makes the *best* leaf into a rule.)
5. A Naïve-Bayes classifier[48] is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. This classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. A naive Bayes classifier considers each of the features to contribute independently of the probability that a sample belongs to a given class, regardless of the presence or absence of the other features.

Table 2 shows the results in terms of training and test errors.

Reviewing the results obtained, we opted for the SVM PEGASOS approach, as SVMs are among the most popular ML techniques due to their capabilities for capturing complex relations between the data without the need for complicated preprocessing.[49] In particular, the PEGASOS approach,[41] that implements the stochastic variant of the PEGASOS method of Shalev-Shwartz et al.[41] was the approach that obtained the best generalization result (together with IB1, a neighboring technique) with the real data set of 48 patients, testing over the MC simulation data set.

After deciding on using the SVM PEGASOS for further experimentation, our next step was to carry out a new

Table 2. Comparison of various classifying approaches applied to the separation of two major heart failure phenotypes.

CLASSIFIER	TRAINING ERROR (%)	TEST ERROR (%)
C4.5	4.17	6.35
Naïve bayes	4.17	9.52
SVM PEGASOS	**2.08**	**4.76**
SMO	16.67	19.05
IB1	**2.08**	**4.76**
OneR	6.25	**4.76**
NNGE	6.25	6.35
PART	4.17	6.35

Note: Bold face is used to highlight best results obtained in training and test errors.

study, aiming at studying the results obtained when patients belonging to the so-called gray zone for EF (ie, the area where 40% < EF < 50%) were included. A set of experiments were carried out, using different cut-off points (EF at 40%, 45%, 50%, 55%) in the training data set (data set 1), to evaluate the consequences of adopting different criteria for defining major HF phenotypes. We also looked at a level of 55%, because it has been suggested that the EF value for women may be higher.[50] As test set, we used data set 3, with implemented physiological constraints to the MC data in terms of EF below 10% or beyond 95%, plus a maximum value for cardiac output. A summary of the results is shown in Table 3. Note that while using different cut-offs, the number of samples belonging to each class will vary.

In detail, the regression analysis results are summarized in Table 4.

Figure 4 shows in the left columns the real labels for the test set data, while in the right it is shown how the samples that belong to the third group (in green color in the figures at the left) are labeled regarding HFpEF and HFrEF. The figure shows both graphs for the 40%, 45%, 50% and 55% cut-offs. It is interesting to see the evolution of the classifications of

Table 3. A summary with the results obtained for EF cut-offs at 40%, 45%, 50%, and 55%.

CUT-OFF FOR EF	40%	45%	50%	55%
Training set (HFpEF/HFrEF)	42/6	40/8	35/13	33/15
Test set (HFpEF/HFrEF/3rd group)	169/94/140	162/108/133	150/137/116	138/168/97
TPR (HFpEF)	1 (169)	0.91 (161)	0.98 (147)	0.99 (136)
TPR (HFrEF)	0.87 (82)	0.96 (104)	0.97 (133)	0.98 (164)
Third group (HFpEF/HFrEF)	133/7	117/16	59/57	74/23

Notes: The rows called TPR (True Positive Rate) contain within parentheses the number of data samples classified as heart failure (HF) class HFpEF (third row) and HFrEF (fourth row), respectively. The third group are the patients needing further classification, ie, with EF >50% and beyond EDVI 97 mL/m^2.

Table 4. A summary with the results obtained for the regression lines with EF cut-offs at 40%, 45%, 50%, and 55%.

CUTOFF (%)	ESVI–EDVI REGRESSION PREDICTED		ESVI–EDVI REGRESSION THIRD GROUP REMOVED	
	SLOPE	INTERCEPT	SLOPE	INTERCEPT
55	HFrEF: 0.69	−9.63	HFrEF: 0.72	−0.97
	HFpEF: 0.33	−6.73	HFpEF: 0.20	+0.95
50	HFrEF: 0.67	−9.24	HFrEF: 0.76	−11.82
	HFpEF: 0.25	−0.82	HFpEF: 0.20	+2.52
45	HFrEF: 0.73	−6.08	HFrEF: 0.74	−7.25
	HFpEF: 0.44	−11.09	HFpEF: 0.28	−0.73
40	HFrEF: 0.80	−9.34	HFrEF: 0.83	−12.56
	HFpEF: 0.48	−12.39	HFpEF: 0.29	+0.06

Note: First column contains slope and intercept for both HFrEF and HFpEF regression lines for predicted labels and the second column contains the same values if the third group is eliminated.

the two groups between the two classes with the different cut-offs.

Figure 5 shows in the right column, the predicted labels for the test set, while the left column shows the same prediction if the samples belonging to the third group are removed, for 40%, 45%, 50%, and 55% cut-offs, so as to be able to see the differences appearing among them. In all cases, regression lines are depicted in green color, and the regression coefficients for both HF classes are added in Table 4.

A. Cut-off 40%.

In Figure 4, we can see the behavior of the third group, which can largely be classified as HFpEF. Interestingly, the separation does not follow the linear division as prescribed by the concept referring to a constant EF value for the cut-off. As can be seen, the points that are labeled differently seem to be located on the border between the main classes. Within this context, it is relevant to emphasize the transition phenomenon where HF patients migrate from HFpEF to HFrEF territory, as well as in the opposite direction. Reportedly, such transitions are observed in 39% of all HF patients (either way) in one study,[17] and 22% in another.[5] This switch of phenotype further complicates the problem of identifying HF patients within clinical subclasses.

B. Cut-off 45%.

Following the same conventions as in the previous cut-off series, and using the information from Table 3, one can derive the number of patients in each class in the data set, and the number of instances in each class of the test set.

As before, the second two graphs from Figure 4 show the behavior of the third group, which can partly be classified as HFpEF. Again, the points that seem to be labeled differently are on the border between the main classes and again suggest a nonlinear division between the two major phenotypes of HF.

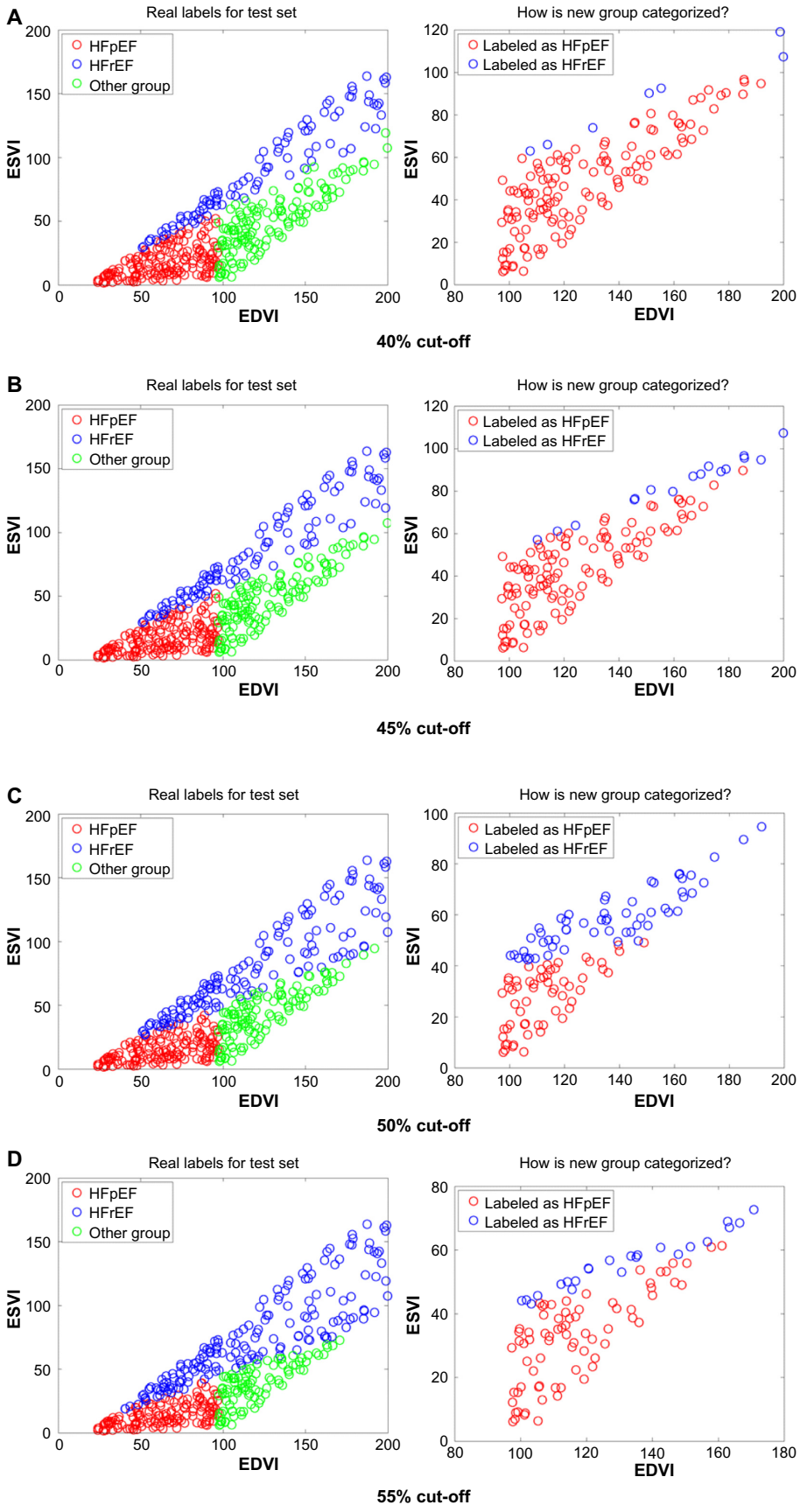

Figure 4. Left column shows the real labels for the test set ($n = 403$, including the third group which is newly assigned to either of the existing phenotypes) for the 40%, 45%, 50%, and 55% cut-off of the patient data set as training test. The right column corresponds to an enlarged picture of the third group data labels and shows in more detail that the algorithm applies a nonlinear division rather than a straight EF line, regardless of the choice at 40%, 45%, or 50% and gradually bends toward the higher iso-EF region.

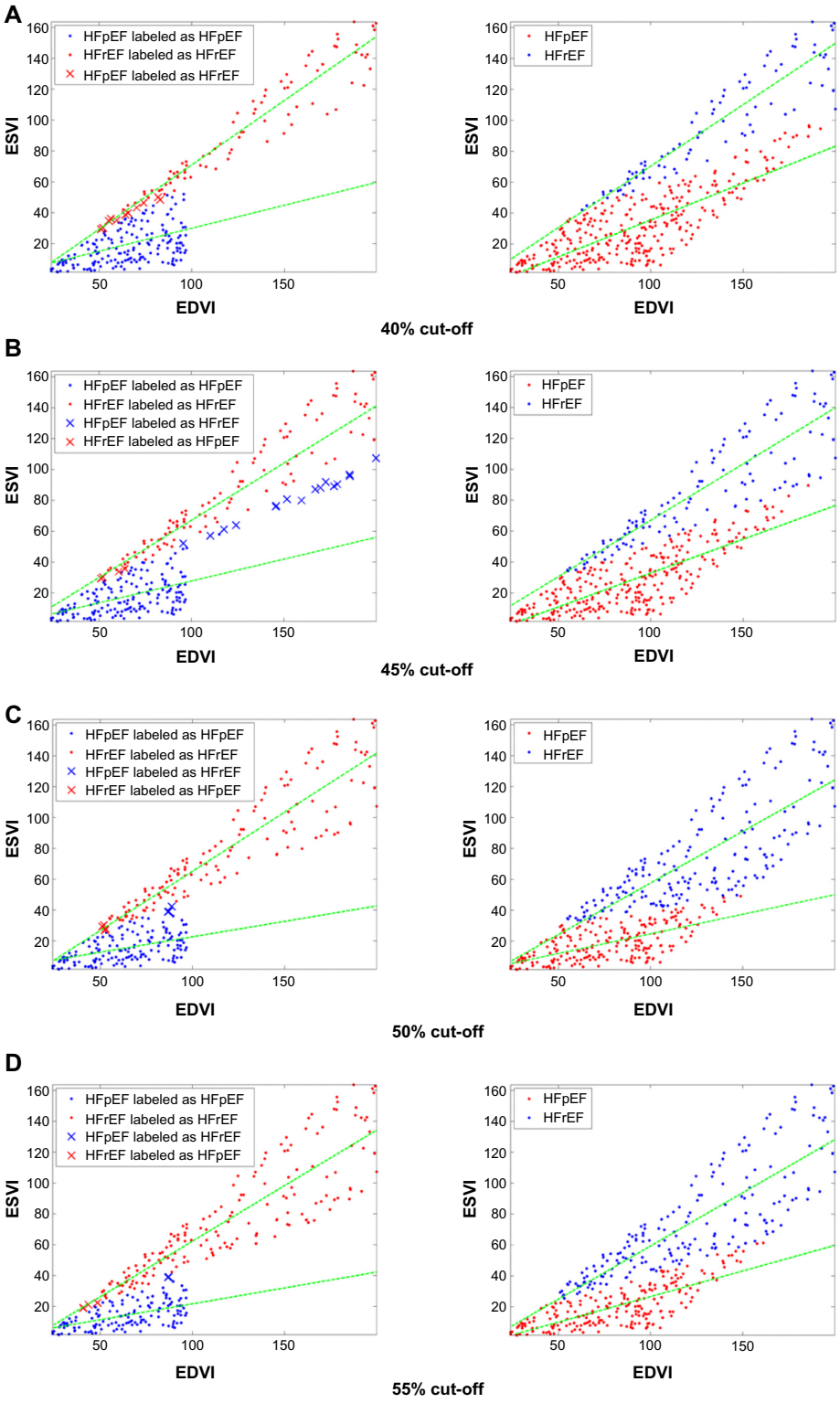

Figure 5. The predicted labels and regression lines for the test set generated with MC for the 40%, 45%, 50%, and 55% cut-offs. This last level could in reality correspond with the suggestion formulated for women, having higher values for EF compared to men.[50] In the right side of the figure, the predictions using the complete test data set is displayed, while on the left side, the samples belonging to the third group are eliminated.

C. Cut-off 50%.

As it can be done with the previous cut-off levels at 40% and 45%, we can also compare the actual labels of the test set and the predicted labels, by examining the third graph in the left column in Figure 4 and the third graph in the right column of Figure 5. Visual inspection of the corresponding graph in Figure 4 reveals that the blue symbols (referring to HFrEF) can quickly be distinguished from the red ones (pertaining to HFpEF) by drawing a straight line almost parallel to the abscissa near ESVI = 40 mL/m^2. Clearly, this is not a perfect

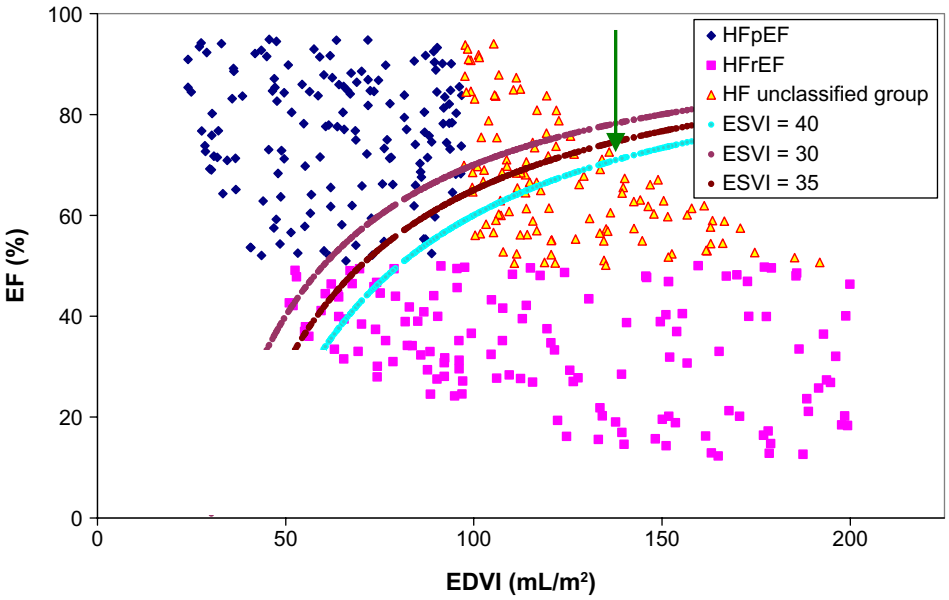

Figure 6. All MC-generated data points ($n = 403$) displayed in the style of the classical paradigm, ie, EF versus EDVI. Two important conclusions can be derived from this graph (see text).

demarcation, but for practical purposes, it can be a convenient rule of thumb, as will later also be illustrated in Figure 6 while following the classical paradigm. As can be seen in Figure 4 in the third row, the number of points of the third group that are now labeled as HFrEF has increased.

D. Cut-off 55%.

Finally, and as above, the corresponding results show that most points belonging to the third group are now again diminishing.

Returning to the old paradigm (illustrated in Fig. 1), we have presented all ($n = 403$) MC-generated surrogate HF patients by means of their data points in Figure 6, which relates EF to EDVI. The group as a whole shows a curvilinear inverse relationship with modest correlation, as previously also documented for a large group ($n = 165$) of patients.[51] The three subgroups are marked by their color-coded symbols. The third group has been introduced to challenge the validity of the linear separation as proposed in published studies by using EF at 40%, 45%, or 50%.[4,11] As seen in Figures 4 and 5, the SVM analysis reveals a clear tendency of the dividing line to bend downward in the higher volume range, thus leaving the concept of a constant EF line, regardless of the specific cut-off level selected for EF. In particular, in the third row in Figure 4 (which refers to the recommended cut-off level for EF at 50%), we observe that a line intersecting near ESVI = 40 mL/m^2 and almost parallel to the abscissa constitutes a reasonable demarcation of the HFpEF and HFrEF territories.

In Figure 6, three curves are inscribed based on constant values for ESVI (at 30, 35, and 40 mL/m^2). Similarly, as shown in Figure 1 for the actual HF patients, we can again separate the HFpEF group (blue lozenges) from the HFrEF group (purple squares) applying a suitable curve referring to ESVI. The third group (yellow-red triangles) represents HF candidates for whom no guidelines have been established yet, ie, with EF >50% and EDVI >97 mL/m^2.[7] Two important conclusions can be derived from Figure 6, both supported by the ML analysis described in this paper:

1. The maximum value for EDVI in this theoretical example is 136 mL/m^2 (see green arrow) in the HFpEF group with cut-off set at EF = 50%. Thus, the previous constraint of 97 mL/m^2 may turn out to be rather superfluous.
2. A fixed value for ESVI (eg, at 35 mL/m^2) reasonably separates HFpEF from HFrEF phenotype. Also, the third as yet unassigned group (triangles) is classified in a rather logical manner.

Discussion

Volumetric components of EF. Figures 1 and 2 illustrate the old (ie, EF versus EDVI) and the new paradigms (namely, the VRG where ESVI is related to EDVI), respectively. It is clear from the definition formula (Eq. 1) that EF depends on its two constitutive components, namely, ESVI and EDVI. The impacts of the two elements are unequal.[51] In fact, ESVI is the major determinant of EF, while often EF and EDVI fail to yield a significant correlation in selected diagnostic groups[13] as also illustrated in Figure 1. Routine cardiac catheterization, as well as application of other appropriate imaging modalities, provides information on EF, yet determining LV size has not achieved the critical role it deserves.[14] The exposé regarding the diagnostic dilemma in congestive HF further points out that everyone accepts that a low EF is abnormal, while unfortunately seldom attention is paid to LV volume itself.[14] In the

present study, we have combined ESVI and EDVI to generate a VRG for the LV. This representation implicitly reveals iso-EF trajectories as linear relationships.[12] Our application of ML methods to classification of subgroups in HF indicates that ESVI may be of pivotal importance. This approach deserves further investigations in longitudinal cohorts, preferably with stratification by gender, age, comorbidities, and therapeutic modalities.[12]

The following sections will critically address several entities that have been instrumental in our approach. Advantages and limitations will be discussed.

Criteria for HF within the old paradigm. HF is a syndrome characterized by a high mortality rate, frequent hospitalizations, reduced quality of life, and a complex therapeutic regimen. Knowledge about HF is accumulating so rapidly that individual clinicians may be unable to readily and adequately synthesize new information into effective strategies of care for patients with this syndrome. Trial data, though valuable, often do not give direction for individual patient management. These characteristics make HF an ideal candidate for practice guidelines.[11] However, criteria formulated for HFpEF may vary considerably.[12] In a recent paper,[52] the guidelines as specified by the ESC and Heart Failure Society of America were interpreted as:

1. Signs and/or symptoms of HF (Framingham or Boston criteria, exercise testing, quality of life questionnaire),
2. EF >50%
3. EDVI <75 mL/m^2
4. Evidence of diastolic LV dysfunction obtained invasively (cardiac catheterization) or noninvasively (transmitral or tissue Doppler or left atrial size), and
5. Exclusion of noncardiac diseases that could cause symptoms commonly present in patients with HF.

Remarkably, not all investigators adhere to the EF >50% requirement; some consistently choose 45% and others 40%.[4,11] Therefore, we have specifically explored all these situations using the SVM tool. Apparently, disagreement exists only on the lower side and never on the higher end (eg, EF at 55%). The troubled range concerns precisely the region known as gray zone. It seems that the epicenter of doubt only pertains to selected patients being positively identified as HFrEF or not. According to the official rules, all patients with 40% < EF < 50% inevitably belong to HFrEF. Yet, many studies find indications to count them collectively as HFpEF. Clinicians are apparently not always sure or disagree about the guideline for HFrEF.

The fact that no study overtly addresses a higher than 50% cut-off is not really what we would expect in view of the notion that women may exhibit a higher EF (cut-off) value.[50] We included this higher value in our study and again detected a nonlinear behavior (Figs. 4 and 5, last row) when searching for a demarcation to distinguish HFrEF from HFpEF. Extrapolating the concept of different EF cut-off values for men versus women may imply that the currently held viewpoint that women form a majority in the HFpEF group has to be reconsidered.

Note also the newly reported EDVI cut-off value of 75 mL/m^2, which is almost 25% smaller than the current ESC standard.[16] In another study, this value is set at <102 mL/m^2.[7] More importantly, many investigations do not pay attention to any boundary condition for EDVI.[4,12]

Measurements of plasma biomarkers such as NT-proBNP are not included in the list given above and only proposed when the diagnosis is not certain. Thus, various research groups neglect the relevance of the EDVI criterion, but the committees do agree concerning an EF cut-off as set at 50%. Surprisingly, investigators do not always respect the almost unanimously chosen value for EF, but show in practice a preference for 40%, 45%, or 50%. [4,12] Clearly defined criteria for diagnosing HFpEF are sorely needed, but evidence validating their use, particularly in the elderly, is lacking.[50] For these reasons, we explored the outcomes of different boundary values for EF. We added in our analysis the 55% cut-off value in order to explore consequences of the suggestion that for women the value may be higher than for men.[50]

VRG and ML models; what can be gleaned?. As a starting point, the EF constraint is applied to both the learning and the testing sets. However, their purposes are different, as well as their approach. EF in the classical paradigm is a linear divider, whereas the distinctions made using ML methods explore a nonlinear hyperplane. Our strategy is to first find an ML method that compares favorably well with the accepted old paradigm. As a next step, we try to gain insight into the decision pattern exhibited by the SVM evaluations. In other words, we take advantage of the observed *errors* in order to be able to detect new and possibly better options for separating diagnostically distinct patient groups. Analyzing the behavior of the ML methods may provide suggestions for alternative classification *rules*. Using this approach, our study advances ESVI as a fascinating candidate, thus supporting the secondary insight derived from Figures 1 and 6.

Our study introduces and combines two important elements for the classification of HF phenotypes: the presentation of HF in the LV volume domain[12,13] and the application of SVM. The volume domain approach overcomes a severe limitation inherent to the use of EF (which solely concentrates on a ratio, thereby largely neglecting the importance of regulation of LV cavity volume). The ML method has the advantage that analysis is not restricted to a linear divider (such as the dichotomy embodied in the use of EF for classification purposes). EF is considered central to the evaluation and management of HF, informs about drug and device therapy, and is widely accepted as a determinant of prognosis.[5] However, over the last 35 years ESVI has increasingly been endowed with a similar pivotal role.[12] Figure 1 is a striking example, showing that the single use of ESVI is almost equivalent to the combined application

of EF and EDVI constraints. Interestingly, the EF and ESVI schools can be fairly well united by recognizing the (nonlinear inverse) relationship between the two indexes.[12,13]

In our HF study, we have included 35 patients with HFpEF and 13 patients with HFrEF.[13] We acknowledge that these sets concern rather limited groups to be used for training purposes when applying ML methods. However, the size of our patient groups exceeds those recruited, eg, for an analysis of the ultrastructural differences[53] when comparing HFpEF (n = 16) and HFrEF (n = 17) in patients, excluding those adherent to beta-blockers.[21] As mentioned before, to compensate for the limited numbers of patients, we applied techniques such as k-fold cross-validation. Also, the MC strategy permits the creation of large groups of surrogate patients suitable for extensive analysis.

The definition of HF has a long history and is still evolving, notably since the identification of various additional subtypes.[12] We applied the strict ESC criteria[16] while using the *gold standard* for measuring LV filling pressure, ie, invasive determination of LV pressure rather than an echocardiographically derived surrogate. Importantly, the clinical data analyzed here were not specifically collected for this study, but were retrospective and primarily part of routine diagnostic investigations. This means that our training set for ML consists of the best possible data available, acquired within an advanced clinical setting. Testing data are generated by MC simulation combined with realistic pathophysiologic constraints.[12] Within this context, it is crucial to emphasize that we are not primarily looking for specific properties of artificially created patients, but rather explore a large set of pertinent data in order to delineate hyperplane criteria to classify subtypes of HF, other than those based upon the somewhat controversial index of EF.[12,14] Application of PEGASOS rendered the highest performance, implying that only few instances were labeled differently from the prediction on the basis of the EF paradigm. However, there exist no documents that prove that EF is the superior criterion. In our study, we employ the EF–EDVI paradigm as a starting point in order to be able to derive further insight from the behavior of ML methods under these circumstances. In that respect, the choice for applying PEGASOS acts as a hardest case scenario for our study, generating familiar outcomes but using totally different routes. That is precisely what we want to perform: to learn from ML methods if alternative approaches can ignite new ways of thinking about the HF classification issue. In this respect, it is remarkable that the ML approach suggests ESVI as a key parameter, while we noted a similar role for ESVI by close inspection of the old paradigm (Fig. 1). It should also be emphasized that ML techniques may employ nonlinear strategies to separate groups. It appears from our evaluations that the ML approach favors a separation which runs somewhat perpendicular to the EF cut-off line. This implies that rather ESVI is invoked as a powerful discriminator.

EF as criterion in HF. The simple metric of EF has been applied for decades to evaluate the status of patients with heart disease.[12–14,18,23,25,26] Particularly, EF has been employed to predict prognosis and document efficacy of surgical and pharmacologic interventions. With the recent recognition of the HFpEF syndrome, however, we are confronted with the fact that EF is not a reliable indicator for *the goodness* of LV performance in millions of HF patients worldwide. In fact, in one study specifically investigating transitions in a cohort of HF patients, it was concluded that *EF is a dynamic factor* related to sex, coexisting conditions, and drug therapy.[5] In an earlier study on HF, we reasoned that under certain conditions EF is related to the arterial–ventricular coupling index, which in turn, is composed of the ratio of two other ratios.[13] Paradoxically, the actual value of EF in some HFpEF patients may even be higher than that in healthy individuals.[12] This notion seems frustrating, but essentially illustrates an overwhelming clinical dilemma: is EF generally useful or is the index only applicable in selected cases?.[12] Admittedly, guidelines indicate that EF is not the only indicator to distinguish HF subtypes and emphasize that EF must be interpreted against the full clinical background of the individual patient.[16] The fact that EF is generally accepted does not by definition imply universal applicability; such appanage would imply a major scientific flaw.[12] In fact, the HFpEF syndrome itself is the best example illustrating that a high (ie, normal or preserved) value for EF apparently does not guarantee something like the presence of a healthy heart, let alone the absence of failure. In addition, the rather arbitrary incorporation of a gray zone (where $40\% < EF < 50\%$) is not acceptable, because that route leaves a major critical group of patients virtually undiagnosed.[12]

The EF–EDVI paradigm versus the VRG. The official ESC criteria formulated for HFpEF imply two boundary conditions, namely, for EF and for EDVI.[16] This concept is illustrated in Figure 1. For the HF patients described in this study, the two HF phenotypes individually yield no significant relationship in the EF–EDVI plane. This is remarkable, because one would rather expect spurious correlations due to the fact that the values on the ordinate are explicitly dependent on the values on the abscissa. Interestingly, Figure 1 also demonstrates that instead of the two criteria formulated by the ESC, a single value of ESVI nicely performs the same task of separating the two phenotypes. Thus, in this case, a fixed value for ESVI suffices as can readily be determined by visual inspection. This notion seems surprising but may easily be appreciated by the observation of near-perfect linear relationships between ESVI and EDVI for selected diagnostic groups,[12] and the inverse nonlinear relationship between EF and ESVI.[12,13,51] These considerations resulted in the model to analyze HF via LV volume regulation by relating ESVI to EDVI.[12,18] Accordingly, the present study employs this novel route to classify HF patients. To this end, we apply ML methods that are relatively new in the field of LV pump function analysis. We applied these analysis methods in

particular to the ESVI–EDVI graph (cf. Fig. 2) because this representation provides highest correlation among all combinations of ESVI, EDVI, EF, and the differential termed SV index.[12,18] Moreover, the combination of ESVI and EDVI has a sound physiological basis, in contrast to (for example) EF versus EDVI (shown in Fig. 1). Another advantage of ESVI versus EDVI is the fact that this presentation allows for the inscription of linear iso-EF lines.[12] Recently, several statistical learning algorithms, including unbiased hierarchical cluster analysis of phenotypic data (with 67 continuous variables) and penalized model-based clustering, have been successfully applied to define and characterize mutually exclusive patient groups making up a novel classification of HFpEF.[37]

The EDVI boundary for HFpEF. The additional boundary condition intrinsic to the current EF-paradigm and concerning LV filling volume (ie, EDVI < 97 mL/m^2) in HFpEF patients seems superfluous. This notion does not come as a surprise, because in practice, most investigators already do not really care about the EDVI constraint. Remarkably, a few studies report constraints for EDVI that differ from the proposed value as formulated above. In one study, a boundary <102 mL/m^2 has been introduced,[7] whereas a recent paper restricts to <75 mL/m^2.[52] Inspection of Figure 1 suggests that the applied ESVI criterion does not preclude the occurrence of any HFpEF data points beyond an EDVI of 97 mL/m^2. Also, the MC simulation–based results indicate that the classification for both phenotypes continues without restrictions for EDVI, although slight nonlinear behavior of the separating curve becomes manifest in the VRG representation. Importantly, we observe again that ESVI may be a relevant candidate to classify the two groups, especially when we introduce the third group (Figs. 4 and 5), which is not assigned to either HFrEF or HFpEF on the basis of current guidelines.[16] As mentioned before, the choice for applying PEGASOS refers to a hardest case scenario for our study. Yet, the SVM approach generated results very similar to the EF paradigm, although obtained via a totally different route. Reviewing alternative ML strategies tested in our study such as K-means and EM, we derive from the pertinent graphs (Fig. 3) that those approaches even more ostensively document dividing lines that are almost perpendicular to the iso-EF line of 50%. In that respect, these outcomes support the strength of the SVM method for classifying HF patients.

Nonlinear divider for HFpEF versus HFrEF. The current paradigm regarding classification of HF phenotypes assumes a linear divider, namely, EF at a particular cut-off level according to the somewhat liberal preference of the individual investigator.[3,4,12] This variability results in the following diversity: EF cut-off at 40% with 22% transitions (both ways in a cohort of 2413 HF patients) within an observation period covering 8 years,[5] at 50% (with 39% transitions either way),[17] at 40% for HFrEF, and at 50% for HFpEF (implying a gray zone between 40% and 50%),[53] 45%–50% but likely higher in women,[50] at 45% (and EDVI <102 mL/m^2, plus evidence of diastolic dysfunction by Doppler echocardiography),[7] at 50%,[30] at 50%,[29] at 45% but without mentioning the EDVI constraint,[6] and at 45%.[8] However, there is no rationale to support linearity. It is very well conceivable that the cut-off value is not a fixed number for EF, but rather varies with gender, age, medication, and the size of the heart. Indeed, the choice of EF at 40%–55% as criterion for HFpEF appears arbitrary as there is evidence that the lower limit for EF in elderly women is much higher.[50] Such a nonlinear discriminator would in particular be more appropriate when dilated hearts are involved. As a matter of fact, the outcomes of the present study also suggest a nonlinear approach, cf. Figures 4 and 5. In contrast, the ML techniques applied here do not assume a linear discriminator, but indeed create nonlinear (hyper) planes to distinguish groups. Adhering to a nonlinear framework also necessitates a re-evaluation of the present simple concept of transition from one phenotype to another, just by crossing a straight dividing line.

The distinction between HFpEF and HFrEF is not always easy in clinical practice, as is sadly illustrated by the frequent introduction of a gray zone for borderline patients.[12] The lack of consensus among investigators regarding the precise cut-off value for EF obviously further complicates the issue. Therefore, the purpose of this study is exploration by advanced techniques to classify HF subtypes and outlining consequences of a particular choice for the EF cut-off when being different from 50%. It should be noted that not all instances classified by SVM when deviating from those assumed on the basis of current guidelines, are necessarily incorrect. The relevance of our findings should not only be derived from the reported level of match and agreement with whatever type of current criteria used in the training set. What is probably a more important outcome that is to discern which alternative path of classification is suggested by the novel approach which is not limited by any restriction regarding a linear divider such as EF. When appreciating the nonlinear approach inherent to ML methods, it is clear that the proposed assignments are rather logical. As a consequence, the actual accuracy of the SVM analysis may in the future turn out to be better than the present outcomes suggest. Also, they make plausible an extension of the HFpEF range toward slightly higher EDVI values than the current boundary of 97 mL/m^2.[12]

ML tools. Computer-assisted methods for the analysis of large databases and the interpretation of their contents have found early application in the medical sciences.[54] Currently, they are employed, eg, for prediction modeling based on electronic health records,[35] data-mining, and machine-learning literature for disease classification and prediction regarding classification of HF subtypes,[36] and more recently, unbiased clustering analysis using dense phenomapping to identify phenotypically distinct HFpEF categories.[37]

Various methods have been devised for classification, example, discriminant analysis[55] and resampling methods.[56] We have selected the SVM approach and reported our findings

based on various testing groups. The search for solid criteria to distinguish various phenotypes of HF is a turbulent field that urgently requires some streamlining. We offer alternative routes (such as SVM) to establish sound criteria based on the VRG to classify patient groups. In general, we have been interested in analyzing what pattern the ML approach is offering for separating groups. This trajectory looks curvilinear in the higher volume range and somewhat inclined to endow a more central role to ESVI. Therefore, the documented curvilinear relationship between EF and ESVI[18,51]strongly suggests that the assumption concerning a linear EF divider may not be of general utility over the complete clinically relevant range.

Further, we paid attention to the evaluation of the consequences of eliminating a gray (borderline) zone (ie, EF between 40% and 50%). Recently, an additional phenotype has been described, featuring recovery of EF after beta-blockade.[12] Further studies are required to ascertain if ML approaches are capable of classifying more than just two phenotypes.

Conclusions

This study proposes an LV volume regulation representation (ie, ESVI in dependence of EDVI), which is shown to have clear advantages over the classical EF–EDVI paradigm. The new representation is combined with SVM tools to classify HF patients. We conclude that ML models offer promise for computer-assisted distinction between the two major phenotypes of HF patients on the basis of ventricular volume data analysis. The description in terms of LV volume regulation deepens our insights regarding subtypes of HF and how to classify them by employing ML models. Our results derived from surrogate MC patients indicate a slight nonlinear behavior of the divider compared to the fixed value implicated in the EF cut-off paradigm. The nonlinearity of EF as divider is supported by our present observation that a fixed value for ESVI offers a reasonable candidate (Figs. 2 and 6), while the relationship between EF and ESVI is intrinsically curvilinear.[51]

Classification primarily guided by LV volumes may have preference over the traditional use of their ratio EF, which latter method apparently invites some investigators to invoke a somewhat artificial borderline (gray) zone. Future application of EF as a cut-off criterion to distinguish HFpEF from HFrEF patients may require incorporation of modifications based on gender, age, and medication. Our analysis method, while using MC-generated surrogate HF patients, suggests an ESVI-associated division pattern. The analysis also indicates that the upper limit value for EDVI, currently imposed by the classical paradigm, may actually be superfluous in HFpEF patients. In addition, selected ML tools combined with the LV volume regulation concept may assist during the classification of individual patients having measurements located in the (clinically often neglected) gray zone where $40\% < EF < 50\%$.

Following Pickering's thoughts[1] regarding the earlier laborious search for an acceptable classification of hypertension using two boundary values, we are ready to accept that in the search for classifying HF phenotypes, we still have a long way to go.

Acknowledgment

The authors appreciate the feedback and constructive suggestions formulated by the three anonymous reviewers.

Author Contributions

Conceived and designed the experiments: PK, GH, AA, VB. Analyzed the data: PK, AA, VB. Wrote the first draft of the manuscript: PK, AA. Contributed to the writing of the manuscript: PK, GH, AA, VB. Collection of clinical data: GH. Agree with manuscript results and conclusions: PK, GH, AA, VB. Jointly developed the structure and arguments for the paper: PK, GH, AA, VB. Made critical revisions and approved final version: PK, GH, AA, VB. All authors reviewed and approved the final manuscript.

REFERENCES

1. Amsterdamska A, Hiddinga A. The analyzed body. In: Cooter E, Pickstone J, eds. *Companion to Medicine in the Twentieth Century*. (Chap. 27). London: Routledge; 2003:420.
2. CDC. *Heart Failure Fact Sheet*. Atlanta, GA: CDC; 2013. [30329–4027 USA].
3. Alba AC, Agoritsas T, Jankowski M, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. *Circ Heart Fail*. 2013;6(5):881–9.
4. Chan MM, Lam CS. How do patients with heart failure with preserved ejection fraction die? *Eur J Heart Fail*. 2013;15(6):604–13.
5. Clarke CL, Grunwald GK, Allen LA, et al. Natural history of left ventricular ejection fraction in patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2013;6:680–6.
6. Gori M, Lam CS, Gupta DK, et al. Sex-specific cardiovascular structure and function in heart failure with preserved ejection fraction. *Eur J Heart Fail*. 2014;16(5):535–42.
7. Kamp O, Metra M, De Keulenaer GW, et al. Effect of the long-term administration of nebivolol on clinical symptoms, exercise capacity and left ventricular function in patients with heart failure and preserved left ventricular ejection fraction: background, aims and design of the ELANDD study. *Clin Res Cardiol*. 2010;99(2):75–82.
8. Shah AM, Shah SJ, Anand IS, et al. Cardiac structure and function in heart failure with preserved ejection fraction: baseline findings from the echocardiographic study of the treatment of preserved cardiac function heart failure with an aldosterone antagonist trial. *Circ Heart Fail*. 2014;7(1):104–115. [Originally published November 18, 2013].
9. Zile MR, Brutsaert DL. New concepts in diastolic dysfunction and diastolic heart failure: part I: diagnosis, prognosis, and measurements of diastolic function. *Circulation*. 2002;105(11):1387–93.
10. Little WC, Zile MR. HFpEF: cardiovascular abnormalities not just comorbidities. *Circ Heart Fail*. 2012;5:669–71.
11. Lindenfeld J, Albert NM, Boehmer JP, et al. HFSA 2010 comprehensive heart failure practice guideline. *J Card Fail*. 2010;16:e1–94.
12. Kerkhof PLM. Characterizing heart failure in the ventricular volume domain. *Clin Med Insights Cardiol*. 2015;9(S1):1–21.
13. Kerkhof PLM, Kresh JY, Li JK-J, Heyndrickx GR. Left ventricular volume regulation in heart failure with preserved ejection fraction. *Physiol Rep*. 2013;1(2):e0007.
14. Buckberg GD. Congestive heart failure: treat the disease, not the symptom – return to normalcy. *J Thorac Cardiovasc Surg*. 2001;121:628–37.
15. Paulus WJ, Tschöpe C, Sanderson JE, et al. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. *Eur Heart J*. 2007;28:2539.
16. McMurray JJ, Adamopoulos S, Anker SD, et al; ESC Committee for Practice Guidelines. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J*. 2012;33(14):1787–847.

17. Dunlay SM, Roger VL, Weston SA, Jiang R, Redfield MM. Longitudinal changes in ejection fraction in heart failure patients with preserved and reduced ejection fraction. *Circ Heart Fail*. 2012;5(6):720–6.
18. Beringer JY, Kerkhof PLM. A unifying representation of ventricular volumetric indexes. *IEEE Trans Biomed Eng*. 1998;45:365–71.
19. Flach P. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. New York, NY: Cambridge University Press; 2014.
20. Shalev-Shwartz S, Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press; 2014.
21. Kerkhof PLM, Baan J, Buis B, Arntzenius AC. Relations between ejection fraction and ventricular volume, and their alteration by chronic beta-blockade. *Br Heart J*. 1981;46:17–22.
22. Yancy CW, Fonarow GC, Albert NM, et al. Influence of age and sex on the delivery of guideline-recommended heart failure care in the outpatient cardiology practice setting: findings from IMPROVE HF. *Am Heart J*. 2009;157:754.e–762.e.
23. Robotham JL, Takata M, Berman M, Harasawa Y. Ejection fraction revisited. *Anesthesiology*. 1991;74:172–83.
24. Manisty CH, Francis DP. Ejection fraction: a measure of desperation? *Heart*. 2008;94:400–1.
25. Marmor A, Jain D, Zaret B. Beyond ejection fraction. *J Nucl Cardiol*. 1994;1:477–86.
26. Hugenholtz PG, Wagner HR. Assessment of myocardial function in congenital heart disease. In: Adams FH, Swan HJC, Hall VE, eds. *Pathophysiology of Congenital Heart Diseases*. Berkeley: University of California Press; 1970:201–30.
27. Kerkhof PLM, Li JK-J, Kresh JY. An analytical expression for the regulation of ventricular volume in the normal and diseased heart. *Cardiovasc Eng*. 2002;2:37–48.
28. Li J, Becher PM, Blankenberg S, Westermann D. Current treatment of heart failure with preserved ejection fraction: should we add life to the remaining years or add years to the remaining life? *Cardiol Res Pract*. 2013;2013:130724. [Published online 2013 October 24].
29. Penicka M, Bartunek J, Trakalova H, et al. Heart failure with preserved ejection fraction in outpatients with unexplained dyspnea: a pressure-volume loop analysis. *J Am Coll Cardiol*. 2010;55:1701–10.
30. Patel HC, Hayward C, di Mario C, Cowie MR, Lyon AR, Rosen SD. Heart failure with preserved ejection fraction: the impact of stricter definitions. *Eur J Heart Fail*. 2014;16(7):767–71.
31. Cristianini N, Shawe-Taylor J. *An Introduction to Support-Vector Machines and Other Kernel-based Learning Methods*. New York, NY: Cambridge University Press; 2000.
32. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: an update. *SIGKDD Explor*. 2009;11(1):10–8.
33. Witten IH, Frank E, Hall M. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Third Edition (The Morgan Kaufmann Series in Data Management Systems), January, 2011.
34. Bramer M. *Principles of Data Mining*. London: Springer-Verlag; 2013.
35. Wu J, Roy J, Stewart WF. Prediction modelling using EHR data: challenges, strategies and a comparison of machine learning approaches. *Med Care*. 2010;48:S106–13.
36. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data mining and machine learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J Clin Epidemiol*. 2013;66:398–407.
37. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131(3):269–79.
38. MacQueen JB. Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symp. On Mathematical Statistics and Probability. Berkeley: University of California Press; 1967:281–97.
39. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Stat Methodol*. 1977;39(1):1–38.
40. Noam Slonim, Nir Friedman, Naftali Tishby. Unsupervised document classification using sequential information maximization. *Proceedings of the 25th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2002:129–36.
41. Shalev-Shwartz S, Singer Y, Srebro N. PEGASOS: Primal Estimated sub-GrAdient SOlver for SVM. *Proceedings of 24th International Conference on Machine Learning*. 2007:807–14.
42. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput*. 2001;13(3):637–49.
43. Aha D, Kibler D. Instance-based learning algorithms. *Mach Learn*. 1991; 6:37–66.
44. Martin B. *Instance-Based Learning: Nearest Neighbor with Generalization* [Ph.D. thesis]. Hamilton, New Zealand; 1995.
45. Holte RC. Very simple classification rules perform well on most commonly used datasets. *Mach Learn*. 1993;11:63–91.
46. Quinlan R. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers; 1995.
47. Frank E, Witten IH. Generating accurate rule sets without global optimization. In: Proceedings of Fifteenth International Conference on Machine Learning; 1998:144–51.
48. John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence. San Mateo; 1995:338–45.
49. Van Gestel T, Suykens JAK, Baesens B, et al. Benchmarking least squares support vector machine classifiers. *Mach Learn*. 2004;54:5–32.
50. Kaila K, Haykowsky MJ, Thompson RB, Paterson DI. Heart failure with preserved ejection fraction in the elderly: scope of the problem. *Heart Fail Rev*. 2012;17(4–5):555–62.
51. Kerkhof PLM, Li JK-J, Kresh JY. End-systolic volume as primary determinant of ejection fraction in HFpEF versus HFrEF. *Europ. J. Heart Fail*. 2013;15 (Suppl. 1): ISSN 1388–98424 (abstract # 1783).
52. Zile MR, Baicu CF, Ikonomidis J, et al. Myocardial stiffness in patients with heart failure and a preserved ejection fraction: contributions of collagen and titin. *Circulation*. 2015.
53. Hamdani N, Paulus WJ, van Heerebeek L, et al. Distinct myocardial effects of beta-blocker therapy in heart failure with normal and reduced left ventricular ejection fraction. *Eur Heart J*. 2009;30(15):1863–72.
54. Kerkhof PLM, Alonso-Betanzos A, Moret-Bonillo V. Medical Artificial Intelligent Systems. *In: Webster's Encyclopedia Electric and Electronic Engineering*. New York: Wiley & Sons; 1999;12:498–508.
55. Agresti A. *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons; 1996.
56. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *Appl Statist*. 1999;48:313–29.