Title: Detection of congestive heart failure from short-term heart rate variability segments using hybrid feature selection approach

Authors: Alan Jovic[a*], Karla Brkic[a], Goran Krstacic[b,c]

Affiliations

[a] University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, HR-10000, Zagreb, Croatia

[b] Institute for Cardiovascular Diseases and Rehabilitation, Draškovićeva 13, HR-10000 Zagreb, Croatia

[c] Josip Juraj Strossmayer University of Osijek, Faculty of Medicine in Osijek, Ulica Cara Hadrijana 10E, HR-31000, Osijek, Croatia

*Corresponding author:

E-mail: alan.jovic@fer.hr

Tel: +385 1 6129 548

Fax: +385 1 6129 653

Abstract

Objectives:

The aim of this work is to investigate the accuracy limits of automated detection of congestive heart failure (CHF) from short-term heart rate variability (HRV) series. Short-term HRV analysis uses 5-minute segments from HRV recordings to diagnose a disorder. This work proposes a hybrid feature selection procedure aimed at finding highly accurate models containing only a few highly informative features, which enables physiological interpretation of the features relevant for the model.

Materials and methods:

Short-term HRV segments are analyzed for CHF diagnosis. Subjects' records from four public PhysioNet databases are considered (66 healthy subjects and 42 CHF subjects). The problem is approached from a machine learning perspective, by extracting 111 linear time domain, frequency domain, time-frequency, nonlinear and symbolic dynamics HRV features. A multistage hybrid feature selection method is proposed that eventually eliminates most features. The method uses a symmetrical uncertainty filter, Naive Bayes wrapper with best first search, and final greedy iterative feature elimination. For classification purposes, we use rotation forest (RTF), radial based support vector machines (SVM), random forest (RF), multilayer perceptron artificial neural network, and *k*-nearest neighbors' classifiers in order to evaluate the feature sets at each step of the process and to obtain as accurate model as possible. Leave-one-subject-out cross-validation evaluation method was used, with two variants: subject-level (coarse-grained) and feature vector-level (fine-grained).

Results:

The results show that the feature selection method is capable of either improving or retaining the classification accuracy of the full feature set (RTF: subject-level ACC = 88.9%, feature vector-level ACC = 85.6%; SVM: subject-level ACC = 89.8%, feature vector-level ACC = 83.5%; RF: subject-level ACC = 87.0%, feature vector-level ACC = 85.5%), while greatly reducing the number of included features, to only four HRV features for RTF and RF, and only two HRV features for SVM.

The resulting best models for subject-level classification achieved are: RTF: ACC = 90.7%, SENS = 78.6%, SPEC = 98.6%, obtained with features: LF/HF ratio, maximum alphabet entropy, alphabet entropy variance, and HaarWaveletSD (scale = 8); SVM: ACC = 88.0%, SENS = 78.6%, SPEC = 93.9%, obtained with features: LF/HF ratio and Rate_U; RF: ACC = 90.7%, SENS = 78.6%, SPEC = 98.6%, obtained with features: LF/HF ratio, maximum alphabet entropy, Rate_U, and Rate_B. Other classifiers provided similar, but somewhat lower results. A comparison of the procedure with the results of individual filter, wrapper, and simple hybrid approaches is provided, which demonstrates the efficiency of the proposed procedure.

Conclusions:

The results suggest that the method can achieve accurate generalizable models for automated diagnosis of CHF from short-term HRV segments in subjects with very few informative features. The choice of the best features and the classification results are similar between the three best classifiers, so the use of any of them with the proposed method is recommended. Nonlinear and symbolic dynamics features are shown to have an important role in the resulting models. The presented methodology may be useful for first-hand screening for CHF as well as for similar diagnostic or automated detection problems in biomedicine.

Keywords

feature selection; feature extraction; heart rate variability; congestive heart failure; classification, electrocardiogram analysis

1 Introduction

Heart rate variability (HRV) analysis measures fluctuations in the lengths of a temporal sequence of cardiac interbeat (RR) intervals. It is used to describe the condition of healthy subjects and patients with various disorders of cardiovascular or other origin, such as myocardial infarction, stable coronary heart disease, congestive heart failure, cardiac arrhythmia, cardiomyopathy, diabetes, renal failure, etc. [1, 2, 11, 35, 36, 65, 69, 74]. HRV time series is traditionally recorded by digital processing of surface electrocardiograms (ECG), mostly from Holter devices [65, 70]. Other techniques are sometimes used for better patient comfort and for fast screening, such as photoplethysmogram [55], impedance plethysmography [73], ballistocardiogram [68], among others.

Disorder modeling using HRV analysis usually includes feature extraction. There are several possible categorizations of physiological time series variability features [1, 70]. HRV analysis traditionally considers linear time domain and frequency domain features to quantify changes in the autonomic nervous system [70]. The influence of some of the time-frequency and nonlinear features was demonstrated in recent literature, especially as a part of feature combinations in HRV-based models of specific disorders [2, 12, 26, 29, 65]. Sassi et al. [65] concluded that, in many instances, nonlinear features seem to provide different information from the traditional features about the complexity of the physiological underlying mechanisms of HRV, which constitutes an important step for the future studies. Faust et al. [12] have shown that spatial filling index and Renyi's entropy of time-frequency distribution were efficient in classification of 8 types of cardiac conditions. Aktaruzzaman et al. [2] have successfully used sample entropy features to improve the accuracy of linear features in sleep stages classification based on HRV. Wang et al. [75] have demonstrated that Poincaré plot nonlinear features and sample entropy are among the best features that show significant difference between healthy subjects and CHF patients groups. Our own previous research indicated that common nonlinear measures improve the results of automatic classification of cardiac disorders in comparison with standard linear time and frequency measures [28, 29, 33]. The use of nonlinear features SD1/SD2, Fano factor and Allan factor improved the classification accuracy of linear feature

combination consisting of 13 time and frequency domain features by 2% on the difficult problem of differentiating between 9 types of heart rhythms [29]. In a second study, we found that adding 10 best ranked alphabet entropy features to a linear feature combination consisting of 10 time and frequency features improved mean sensitivity of six types of arrhythmia classification by about 2% [33]. These results support the use of nonlinear measures in feature combinations for cardiac disorder classification.

Congestive heart failure (CHF) is a serious medical condition of the heart with high prevalence and progressive worsening. The prognosis for a diagnosed CHF patient depends on a number of contributing demographic factors and co-morbidities [61]. It is very important that the disease is diagnosed early so that an appropriate treatment, usually with a combination of medications can be started as soon as possible [17, 73]. Although arriving at certain diagnosis for CHF is an elaborate process involving detailed signs and symptoms evaluation, echocardiography and natriuretic peptides measurement [72], there have been many attempts described in literature to automatically diagnose CHF using HRV features alone [28, 40, 45, 53, 67, 71, 78]. The application of HRV analysis in clinical practice has, however, remained limited [59].

Using HRV, CHF diagnosis is traditionally considered using long-term (usually 24h) recordings, the analysis of which could lead to accurate results [40, 44, 45, 46, 67, 71]. Nevertheless, automated diagnosis from long-term recordings may have limited applicability in practice. The approach supposes that the patient was continuously monitored for 24h, usually with Holter ECG monitoring device or in a hospital setting. Recent research considers using short-term (5-minute) HRV analysis for detection of CHF. The basis for such a reasoning stems from several important studies. Perkiömäki et al. [57] showed that both traditional linear and nonlinear features measured on long- and short-term scales compare relatively good with respect to Pearson's correlation coefficient (r = 0.56 on average for healthy subjects at rest). Maestri et al. [42] showed that some of the short-term HRV temporal scale variables have independent prognostic power for cardiac death and urgent transplantation outcomes in CHF. Also, detrended fluctuation analysis factor for short-term HRV, DFA α_1, was found

to be significant as an independent prognostic marker of sudden cardiac death [24] and ventricular tachyarrhythmias [58]. Therefore, although CHF diagnosis from continuous long-term recordings may be more accurate, the focus of research has recently switched to low-cost, non-invasive, and portable classification methods that are based on short-term HRV analysis and thus may provide early screening for the disorder [18, 26, 40, 53, 55].

The aim of this work is to investigate the accuracy limits of models for automated diagnosis of CHF from short-term (5-minute) HRV segments. Accurate models for discerning CHF from healthy state based on short-term HRV could be clinically relevant, because such a method could lead to earlier administration of patients to hospitals, earlier treatment, and prevention of early death resulting from CHF. However, our current study does not consider early detection of CHF with all the clinical implications, but rather focuses on automated detection of CHF from short-term HRV segments only. In order to obtain the best predictive model, we approach the problem from a machine learning prospective. First, a large number of traditional linear (time and frequency) and recent nonlinear features from subjects' records from several public databases [16] are obtained. Next, several feature selection approaches are considered, which eventually eliminate features that are irrelevant or redundant in achieving the optimal model for CHF detection. The best feature selection approach, a hybrid one, is proposed. Unlike other simpler feature selection approaches used in literature [7, 26, 27, 50, 79] that reduce the feature set size only moderately, the complex hybrid feature selection used in this work succeeds in obtaining a very small set of highly informative HRV features, which allows us to discuss about the physiological interpretation of the features that consitute the best models. The proposed feature selection method is comprehensive, while retaining reasonable computational efficiency when compared to exhaustive search. By employing leave-one-subject-out cross-validation, the models are built using several classification algorithms: rotation forest [64], support vector machines [60], random forest [6], multilayer perceptron [21], and *k*-nearest neighbors [48] in order to ensure high accuracy.

2 Materials and methods

2.1 Dataset

In this work, we use records from four PhysioNet databases [16], as shown in Table 1. Two databases contain records from healthy subjects with mostly normal heart rhythm (NSR) and two databases from

Table 1 The PhysioNet databases used in the study

Database name / short	Type	Sampling frequency, Hz	Records	Records used	Omitted records	Segments used
Normal sinus rhythm RR interval database / NSR1	NSR	128	54	49	nsr032, nsr033, nsr039, nsr043, nsr045	1793
MIT-BIH Normal sinus rhythm database / NSR2	NSR	128	18	17	19088	713
Congestive heart failure RR interval database / CHF1	CHF	128	29	27	chf 205, chf 207	886
BIDMC Congestive heart failure database / CHF2	CHF	250	15	15	-	690

congestive heart failure (CHF) patients. All the records in the databases were annotated with R-peaks' occurrence times and beat types, which was performed either with automated analysis only (NSR2 and CHF2 databases) or with automated analysis and manual corrections from experts (NSR1 and CHF1). In this work, we assume that the given R-peak times are correct and do not propose any specific R-peak detection algorithm.

NSR1 database includes beat annotation files for 54 long-term ECG recordings of subjects in normal sinus rhythm (30 men, aged 28.5 to 76 (mean 67.0)) and 24 women, aged 58 to 73 (mean 65.0)). NSR2 database includes 18 long-term ECG recordings from patients with no significant arrhythmias (5 men, aged 26 to 45 (mean 35.8) and 13 women, aged 20 to 50 (mean 33.8)). CHF1 database includes beat annotation files for 29 long-term ECG recordings of subjects aged 34 to 79 (mean 55.3), with congestive heart failure (NYHA classes I, II, and III). Subjects included 8 men and 2 women; gender is not known for the remaining 19 subjects. CHF2 database includes long-term ECG recordings from 15 subjects (11 men, aged 22 to 71 (mean 54.7), 4 women, aged 54 to 63 (mean 59.3)) with

severe congestive heart failure (NYHA class III–IV). The subjects' diversity in the analyzed databases contributes to a broader capacity for detecting CHF from HRV signal in general population than most of the related work [26, 40, 50, 53].

First 120 minutes of each record is considered in the analysis, with a 50% overlap between the (short-term) 5-minute segments. Some records were omitted prior to the analysis as a result of detailed manual inspection. The records omitted from NSR1 and CHF1 databases had less than 10 valid (no recording pauses longer than 3 s, no large number of arrhythmic beats or recording artifacts) 5-minute segments (out of 47 possible segments). A record from NSR2 was omitted, because of unreliable detection of R-peaks. The records from CHF1 and CHF2 databases were from CHF patients having mostly normal heart rhythm with occasional ectopic beats.

All segments in the first 120 minutes of each record with less than 90% of normal beats were disregarded. The ectopic beats (mostly premature ventricular or atrial contraction) in the remaining segments were interpolated using linear interpolation, which is a recommended way for removing occasional ectopic beats [54]. The final dataset consisted of 4082 segments from 108 subjects. There were 2506 segments from 66 healthy subjects with NSR and 1576 segments from 42 subjects with CHF.

2.2 Features

Standard time domain and frequency domain features for short-term HRV were extracted from 5-minute segments, as suggested by HRV guidelines [65, 70], Table 2. We added the linear time domain pNN20 feature due to inconclusive discussion about whether pNN50 or pNN20 should be used for HRV analysis [25]. The nonlinear features mentioned as promising in the recent HRV guidelines [65]: α (1/f power law exponent), fractal (Hausdorff) dimension D, correlation dimension D_2, Hurst exponent, the largest Lyapunov exponent, deceleration/acceleration capacity, and heart rate turbulence features (e.g. onset and slope) were not considered in this work because they require a large number of RR intervals for accurate estimation (24h records or more) [5, 24, 59, 65].

Table 2 HRV features used in the study

Category	Feature	Description	No. of features	Reference
Linear, time-domain, statistical or geometric features	AVNN	Average of all NN intervals	1	[65]
	SDNN	Standard deviation of all NN intervals	1	[65]
	RMSSD	The square root of the mean of the sum of the squares of differences between adjacent NN intervals	1	[65]
	pNN50	Percentage of the number of pairs of adjacent NN intervals differing by more than 50 ms	1	[65]
	pNN20	Percentage of the number of pairs of adjacent NN intervals differing by more than 20 ms	1	[25]
	SDSD	Standard deviation of differences between adjacent NN intervals	1	[65]
	HRV_TI	Total number of all NN intervals divided by the height of the histogram of all NN intervals measured on a discrete scale with bins of 1/128 s	1	[65]
	TINN	Baseline width of the minimum square difference triangular interpolation of the highest peak of the histogram of all NN intervals	1	[65]
	CVrr	Coefficient of variation of RR intervals	1	[51]
Frequency-domain*	Total PWR	Spectral power in the range $f \leq 0.4$ Hz	1	[65]
	VLF	Spectral power in VLF range ($f \leq 0.04$ Hz)	1	[65]
	LF	Spectral power in LF range ($0.04 \leq f \leq 0.15$ Hz)	1	[65]
	HF	Spectral power in HF range ($0.15 \leq f \leq 0.4$ Hz)	1	[65]
	LF/HF	Ratio LF / HF	1	[65]
	SpectEn	Spectral Shannon's entropy	1	[3]
Time-frequency	HaarWavSD	Standard deviation of Haar 's wavelet for three scales (sc3, sc4, sc8)	3	[71]
Nonlinear, phase space	SD1/SD2	Poincare plot standard deviations ratio	1	[65]
	SFI	Spatial filling index, d = 2, bins = 5	1	[28]
	CTM	Central tendency measure	1	[10]
	STA1, STA2	Sequential trend analysis	2	[3]
	REC, LMean, DET, RecShEn, Lam	Recurrence plot features, d = 2, lag = 1	5	[1]
Nonlinear, fractal	DFA α_1, DFA α_2	Detrended fluctuation analysis short-term (<=11 beats) and mid-term (>11 beats) complexity	2	[65]
Nonlinear, entropy	CCE	Corrected conditional Shannon's entropy, bins = 5	1	[62]
	RenyiEn	Renyi's entropy, order = 3	1	[76]
	SampEn1 - SampEn4, MaxSampEn	Sample entropy, $SampEn_1$ - $SampEn_4$: r = { 0.1, 0.15, 0.2, 0.25} σ, MaxSampEn: r = {0.01 - 0.5 σ, step = 0.01}, m = 2	5	[63]
	FuzzyApEn1 - FuzzyApEn4, MaxFuzzyApEn	Fuzzy approximate entropy, $FuzzyApEn_1$ - $FuzzyApEn_4$: r = { 0.1, 0.15, 0.2, 0.25} σ, MaxFuzzyApEn: r = {0.01 - 0.5 σ, step = 0.01}, m = 2	5	[77]
	MultiSampEn, scales = 2..10	Multiscale sample entropy, scales = 2..10	9	[22]
Nonlinear, other	LZComp	Lempel-Ziv complexity	1	[81]
	$MultiA_IMean$, $MultiA_IStDev$	Multiscale asymmetry index mean and standard deviation, scale = 10	2	[9]
	Allan factor	Allan factor, scale = 10	1	[29]
Symbolic dynamics	AlphEnAver, MaxAlphEn, AlphEnVar, Rate_A - Rate_AA,	Alphabet entropy average, maximum, and variance; rate of occurrence for all letters; alphabet entropy averages for all letters; threshold for no-change: Θ = 20 ms	56	[33]

	AverAlphEn_B - AverAlphEn_AA

*Lomb-Scargle PSD estimate

The nonlinear entropy measures, such as SampEn, CCE, FuzzyApEn and others were included in the study as they were proposed to quantify the entropy rate of short- to mid-length HRV [65]. Other promising methods for short-term HRV analysis were also included, such as spatial filling index, central tendency measure, sequential trend analysis, recurrence plot, Lempel-Ziv complexity, Allan factor, multiscale asymmetry index, etc. Alphabet entropy (AlphEn), our recently introduced symbolic dynamics method that measures both qualitative and quantitative signal information for short-term HRV analysis was also included [33]. Although we acknowledge the role of gender, age, fitness and other anthropometric factors in diagnosis and prognosis of CHF [72], we opted not to consider them in the model, because such data may be unavailable, and because we wanted to model only the differences between healthy subjects and CHF patients within the heart rhythm itself. A total of 111 HRV features were considered in this work.

2.3 Feature selection and classifier methods

Feature selection methods aim at finding and removing irrelevant HRV features, which are the features that do not contribute to the quality of the model. Also, some of these methods may also remove redundant features, which are the features that do not contribute to the model if some other features are included. Having a small number of relevant features significantly accelerates the required feature extraction, model construction, and may lead to more accurate results [23].

Feature selection methods that suppose feature independency or near-independency can be categorized into filter, wrapper, embedded, and hybrid approaches [31]. Filter methods select a subset of features without using a learning algorithm. Wrapper methods use a learning algorithm to evaluate the results produced by the selected features in classification. Embedded methods perform feature selection during the process of training, which is specific to the applied learning algorithm (e.g. random forest

algorithm uses an out-of-bag set to rank features). We do not consider embedded methods in this work. Hybrid methods exploit the advantages of filter and wrapper approaches by first using a filter approach to eliminate many irrelevant or redundant features and then running the slower wrapper method to obtain the final feature set [23].

In our work, we employ four procedures that use filter, wrapper, and hybrid approaches on the extracted HRV features (Table 2) and comparatively analyze the resulting classification performance (NSR vs. CHF). We examine the filter measure called symmetrical uncertainty (SymUnc) [80], the wrapper method using Naive Bayes (NB) classifier [15], the hybrid method that uses SymUnc and NB, and lastly, the hybrid method that uses SymUnc, NB, and final iterative greedy feature set reduction. For classification of the reduced feature sets, we use the rotation forest [64], random forest [6], multilayer perceptron artificial neural network [21], radial kernel-based support vector machine [60], and *k*-nearest neighbors algorithms [48]. In continuation, we briefly describe these methods. More details can be found in corresponding literature.

2.3.1 Filter: symmetrical uncertainty

SymUnc is a filter method that ranks the quality of a feature using the expression [80]:

$$SymUnc = 2\frac{H(C) - H(C \mid F)}{H(C) + H(F)} \quad (1),$$

where C is the goal (class) variable, F is the evaluated feature, $H(C)$ is Shannon's entropy of the class variable, $H(F)$ is Shannon's entropy of the evaluated feature, and $H(C \mid F)$ is conditional Shannon's entropy of the class variable C given the feature F. SymUnc compensates for the information gain's ($H(C)$ - $H(C \mid F)$) bias toward features with more values and normalizes its values to the range [0,1], with the value 1 indicating that knowledge of the value of either feature completely predicts the value of the other and the value 0 indicating that feature and class variables are independent. The features are ranked in descending order of SymUnc. Classification accuracy is evaluated for the ranked feature

subsets with successive 10% decrease in the number of features retained in the set (starting from 100%).

We also tried other filter-based approaches: information gain, gain ratio, chi square, and OneRule [18]. However, SymUnc provided the best results in terms of classification accuracy of the selected subset of the first N ranked features. A filter method called uFilter, which is an improvement over Mann-Whitney U-test for feature selection, was not inspected as it does not show clear superiority over information gain and other filters but may be considered in future work [56].

2.3.2 Wrapper: Naive Bayes and 'best first' search

Naive Bayes (NB) algorithm can be used as a fast approach for wrapper-based discovery of a subset of important features. NB is a specific type of Bayesian classifier that considers all predictor features independent of one another, given the target class [15]. Existence of highly-correlated attributes in a dataset can influence the learning process and reduce the number of successful predictions. Additionally, NB assumes normal distribution for predictor attributes. NB classifier outputs posterior probabilities as a result of the classification procedure. These probabilities are used to evaluate feature subsets that were used for learning the classifier. We used the 'best first' forward direction search strategy for obtaining the feature subsets. This strategy uses greedy hill climbing with backtracking capabilities. It starts from an empty feature subset and inspects how the addition of a feature to the set influences the output of the NB classifier. The feature that increases the accuracy of NB the most is kept in the selected set. When backtracking to a smaller set and examining all such paths through the feature set does not lead to better results, the search ends.

2.3.3 Hybrid #1: symmetrical uncertainty filter and Naive Bayes wrapper

A novel hybrid feature selection process, which we denote as Hybrid #1, is proposed that uses both the SymUnc filter and NB wrapper approach. First, SymUnc is used to retain the optimal number of features L_{OPT} that does not reduce classification accuracy of the classifier. Then, on the reduced set, NB wrapper is used to further reduce the feature set to $K_{NB} < L_{OPT}$. In this way, the time needed for

using the wrapper on the set of all features is significantly shortened, while retaining the feature set with high accuracy selected by SymUnc.

2.3.4 Hybrid #2: symmetrical uncertainty, Naive Bayes wrapper and final greedy search

Hybrid #1 obtains a small set of features. While a further decrease in the number of features may lead to diminished accuracy, it may still be possible to reduce the set and retain high accuracy. The final reduction may be done in two ways: exhaustive and non-exhaustive [39]. Since it is not guaranteed that Hybrid #1 leads to a feature set small enough for exhaustive search, in which all feature subsets are examined, we propose a greedy search for the final reduction. As a criterion for the greedy search, the feature whose removal leads to the smallest drop (or perhaps a largest increase) in the classifier's accuracy for the subject-level evaluation threshold of $\Theta_E = 50\%$ (see Section 2.4) is selected for elimination. The procedure is iteratively repeated until a significant drop in accuracy with respect to the previous feature set is observed. Note that, in some cases, dropping even a single feature from Hybrid #1 set may result in weaker performance. This process, which we denote as Hybrid #2, leads to $J_{GIE} \leq K_{NB}$ features. The Hybrid feature selection processes #1 and #2 are depicted in Fig. 1.

2.3.5 Rotation forest

Rotation forest (RTF) is a highly efficient and accurate decision tree ensemble classifier, like random forest [6], proposed by Rodriguez et al. [64]. In most applications, the C4.5 decision tree algorithm is used as the base learner. The algorithm focuses on presenting transformed data to the classifier by using a projection filter, most commonly, the principal component analysis (PCA) filter [37]. Let the number of base classifiers in rotation forest be K. In order to create the training set for each base classifier, the instances are first sampled using the bootstrap method. Next, the feature set is randomly split into M groups and PCA is applied to each group. All the eigenvectors are retained as new features in order to preserve the variance in the data. The idea why these M data transformations are performed is to encourage simultaneously the individual accuracy and the diversity of the classifiers within the ensemble. M is usually set to 3 but may be changed with mostly insignificant impact on the results.

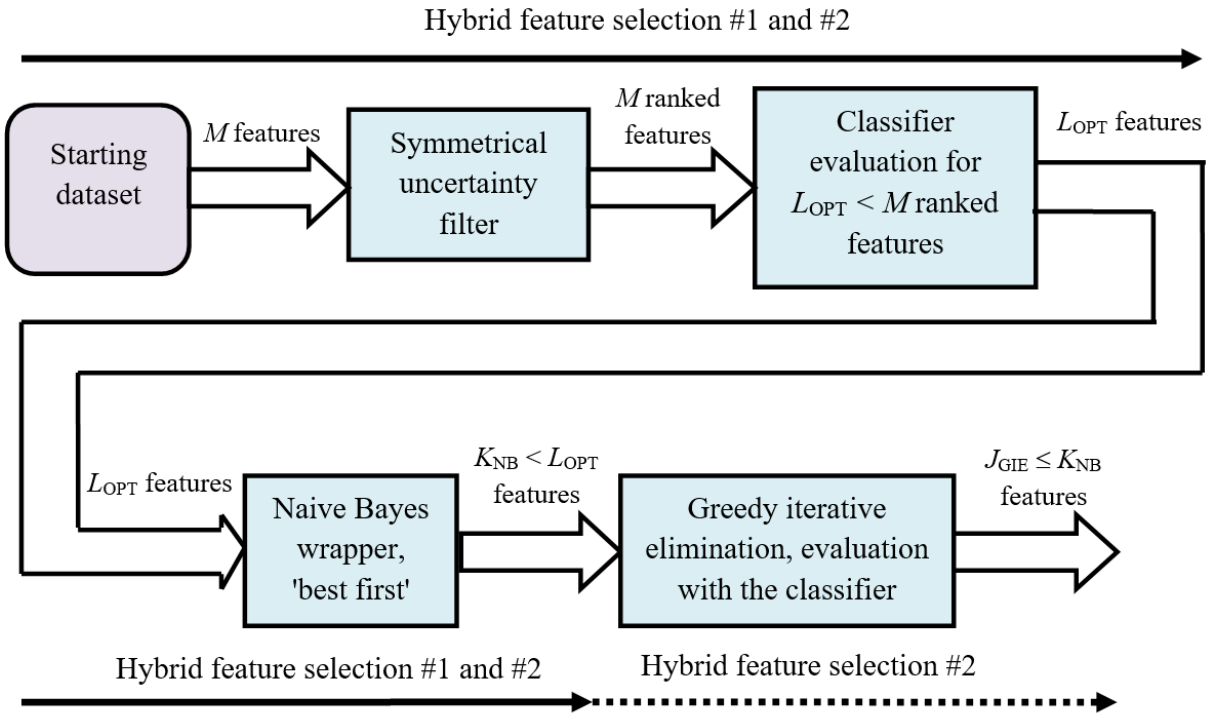

Fig. 1 The Hybrid feature selection processes #1 and #2 proposed in this work

Diversity is achieved through random splitting of the feature set, and accuracy is sought by retaining all the principal components. The only significant hyperparameter of rotation forest is the number of bootstrap iterations (i.e. the size of the decision forest) [37]. We varied the size of the forest in range {20, 50, 100, 200}.

2.3.6 Radial kernel-based support vector machine

Support vector machine (SVM) is a kernel-based machine learning family of methods that are used to accurately classify both linearly separable and linearly inseparable data. The basic idea, when the data are not linearly separable, is to transform them to a higher dimensional space by using a transformation kernel function. In this new space, the samples can usually be classified with higher accuracy. Many types of kernel functions have been developed, with the most used ones being polynomial and radial-basis. Sequential minimal optimization (SMO) algorithm resolves quadratic programming optimization problem that arises when determining the maximum margin hyperplane of the support vector machines classifier [60]. This is a standard SVM learning algorithm. The algorithm

is parametric and deterministic. Two hyperparameters of the SVM classifier were varied: cost parameter C in range {1, 2, 5, 10, 20} and kernel exponent γ in range {0.05, 0.1, 0.15, 0.2, 0.3, 0.4}.

2.3.7 Random forest

Random forest (RF) is an ensemble decision tree learner developed by Breiman [6]. Decision trees choose their splitting attributes (in our case, the extracted HRV features) from a random subset of k attributes at each internal node. The best split is taken among these randomly chosen attributes and the trees are built without pruning. Random forest avoids overfitting due to two sources of randomness - random attribute subset selection and bootstrap data sampling. Breiman has shown that if one constructs the forest consisting of a high-enough number of random trees, the overall classification error will be minimized, and the accuracy will reach a plateau [6]. The interpretation of the ensemble's results is, however, problematic due to the inability to extract useful rules out of the forest. RF is parametric with respect to the number of trees in the forest, but also a stochastic algorithm because of its two sources of randomness. Therefore, the algorithm must be run several times in order to obtain reliable estimates of mean value and variance of the evaluation measures. In our work, we report on the mean value of 5 algorithm runs. We varied the number of trees in the forest in range {20, 50, 100, 200}.

2.3.8 Multilayer perceptron artificial neural network

A feed-forward multilayer perceptron artificial neural network (MLP) is comprised of an input layer, one or more hidden layers of artificial neurons and an output classification layer. The network learns by adapting the weights of the connections between the neurons of consecutive layers, e.g. see [21]. Artificial neural network is considered as a black box that gives classification predictions with no explanation as to how it works, excluding some very specifically designed networks. The main advantages of the use of MLP are the efficiency of the classifier and resistance to outliers. The hyperparameters used for training the MLP in Weka were varied: learning rate in range {0.4, 0.45, 0.5} (for weights adaption) and momentum in range {0.15, 0.2, 0.25, 0.3} (applied to weights). 1000

learning epochs were used, with the standard architecture of one hidden layer with the number of nodes equal to (no. of features + no. of classes) / 2.

2.3.9 K-nearest neighbors

K-nearest neighbors (*k*-NN) is the basic non-parametric instance-based learning method. The classifier has no training phase; it just stores the training set samples. In the test phase, the classifier assigns a class to an instance by determining the *k* instances that are the closest to it, with respect to Euclidean distance metric. A class is assigned as the most commonly occurring one among the *k*-nearest neighbors of the test instance. This procedure is repeated for all test set instances. The simple method is highly useful and usually quite accurate in practice, especially in the cases when there are many more instances than the number of attributes and in the presence of noise (for $k > 1$) [48]. In our work, we varied *k* in range {3, 5, 7}.

2.4 Evaluation procedure details

The classifiers are evaluated using leave-one-subject-out cross-validation (LOSOCV) in two variants. First, feature vector-level classification of segments is used. Here, the classifier is trained on all segments from all subjects, except on the segments from a single subject used for testing. The testing results are recorded, and the procedure is repeated so that all segments from all subjects are tested. The results from this fine-grained approach may be overly pessimistic, because here it is assumed that we do not have any data available on the new subject. This approach is different from the one proposed by Park and Kang [52], where the classifier was cross-validated on segments from only a single subject (personalized classifier).

In the second evaluation scenario, which we call subject-level classification, we also train the classifier on all segments from all subjects except for the tested one. The difference is that, for the tested subject with *N* segments, we select an evaluation threshold Θ_E for classification (coarse-graining), which would identify the subject as having CHF. If $N_{CHF} / N_{TOT} > \Theta_E$, where N_{CHF} is the number of segments

classified as CHF, and N_{TOT} is the total number of segments for the subject, then the subject is classified as CHF.

Depending on the correct label for the segment (feature vector-level) or subject (subject-level), which is NSR or CHF, the classification is counted as true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In order to be comparable with most of the other research in the field, we report on the classification accuracy (ACC), sensitivity (SENS) and specificity (SPEC) of the constructed model, given as [20]:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$SENS = \frac{TP}{TP + FN} \quad (2).$$

$$SPEC = \frac{TN}{TN + FP}$$

Feature selection in our work starts either with SymUnc filter or NB wrapper. For application of both feature selection methods, we consider only the training sets, consisting of feature vectors from all subjects, except the one used for testing (LOSOCV). The number of such training sets equals the number of subjects, denoted as *N*. Note that SymUnc may obtain slightly different feature rankings for each training set. For the classifier evaluation on the *N* test sets (individual subjects), we take the first L_{OPT} features, such that for all *N* feature set rankings obtained on the training sets, these L_{OPT} features are most commonly the highest-ranking ones. Then, for Hybrid #1 method, NB is used on these L_{OPT} features to obtain K_{NB} features, see Fig. 1. When analyzing data with NB wrapper, the method may obtain *N* slightly different feature subsets (no ranking is done with NB). For the classifier evaluation, in this case, we select the average number *M* of features appearing in the *N* feature sets, and the most commonly appearing *M* features are included.

2.5 Software used for the analysis

For records preprocessing and feature extraction, we use HRVFrame framework [30, 32], v2.17, which is openly available under LGPL v3.0 license. Feature selection and classification is performed using Weka platform [20], v3.8. The implementation of the novel feature selection method itself and the implementation of feature vector-level and subject-level classification are done in Python, relying on Weka's feature selection methods, classifiers and (in part) evaluation methods.

3 Results

In Table 3, we summarize the best results for the analyzed dataset in the case where no feature selection was performed and the classifiers were run on the set of all features, with the evaluation procedure specified in Section 2.4. For subject-level classification, the best achieved accuracy (ACC) of 89.8% is obtained by SVM for the evaluation threshold $\Theta_E = 50\%$, which is highlighted in the table. For feature vector-level classification, the best achieved accuracy (ACC) of 85.6% is obtained by RTF. The best results were obtained when considering: the rotation forest consisting of 50 trees, SVM

Table 3 Dataset classification results using several classifiers (ACC/SENS/SPEC), %, no feature selection

Classification	RF	RTF	MLP	SVM	*k*-NN
Feature vector-level classification	85.5/72.6/93.6	85.6/72.6/93.8	82.1/74.7/86.7	83.5/73.9/89.5	81.3/69.0/89.0
Subject-level classification					
$\Theta_E = 10\%$	80.6/88.1/75.8	81.5/90.5/75.8	69.4/92.9/54.5	73.1/92.9/60.6	73.1/90.5/62.1
$\Theta_E = 20\%$	86.1/83.3/87.9	87.0/85.7/87.9	78.7/85.7/74.2	81.5/88.1/77.3	81.5/83.3/90.3
$\Theta_E = 30\%$	88.0/81.0/92.4	88.0/78.6/93.9	81.5/83.3/80.3	86.1/83.3/87.9	86.1/81.0/89.4
$\Theta_E = 40\%$	85.2/69.0/95.5	88.0/73.8/97.0	85.2/78.6/89.4	89.8/83.3/93.9	86.1/78.6/90.9
$\Theta_E = 50\%$	87.0/69.0/98.5	88.9/73.8/98.5	88.0/76.2/95.5	89.8/83.3/93.9	88.0/71.4/98.5
$\Theta_E = 60\%$	87.0/69.0/98.5	87.0/66.7/100.0	88.0/73.8/97.0	88.0/71.4/98.5	86.1/64.3/100.0
$\Theta_E = 70\%$	85.2/64.3/98.5	85.2/61.9/100.0	86.1/64.3/100.0	86.1/64.3/100.0	85.2/61.9/100.0
$\Theta_E = 80\%$	84.3/59.5/100.0	83.3/57.1/100.0	84.3/59.5/100.0	82.4/54.8/100.0	75.9/38.1/100.0
$\Theta_E = 90\%$	80.6/50.0/100.0	79.6/47.6/100.0	77.8/42.9/100.0	76.9/40.5/100.0	74.1/33.3/100.0
$\Theta_E = 100\%$	76.9/40.5/100.0	75.9/38.1/100.0	67.6/16.7/100.0	72.2/28.6/100.0	67.6/16.7/100.0

having the cost parameter C = 10 and kernel exponent γ = 0.3, random forest consisting of 50 trees, MLP with the learning rate = 0.45 and momentum = 0.2, and k = 3 for k-NN classifier. From Table 3, it can be seen that the best achieved accuracy for subject-level classification obtained for Θ_E = 50% or Θ_E = 60% differs by no more than 2% among the classifiers, but still, RTF and SVM achieve the overall best results. Note that the optimal SENS of more than 90% is gained for a low evaluation threshold of 10%. However, at the same threshold, SPEC is at its lowest value (from 55% to 76%, depending on the classifier), which means that about 1/2 to 1/4 subjects are wrongly classified as CHF, which is unacceptable in practice. On the other hand, the threshold of roughly 60–70% or more guaranties perfect SPEC, but at the same time, only roughly 2/3 of the subjects with CHF are correctly identified (SENS) in such a case. In Fig. 2a–2e, we present ROC curves (1 - SPEC on the x-axis vs SENS on the y-axis) [20] of different threshold levels (10–100%) for some of the variations in

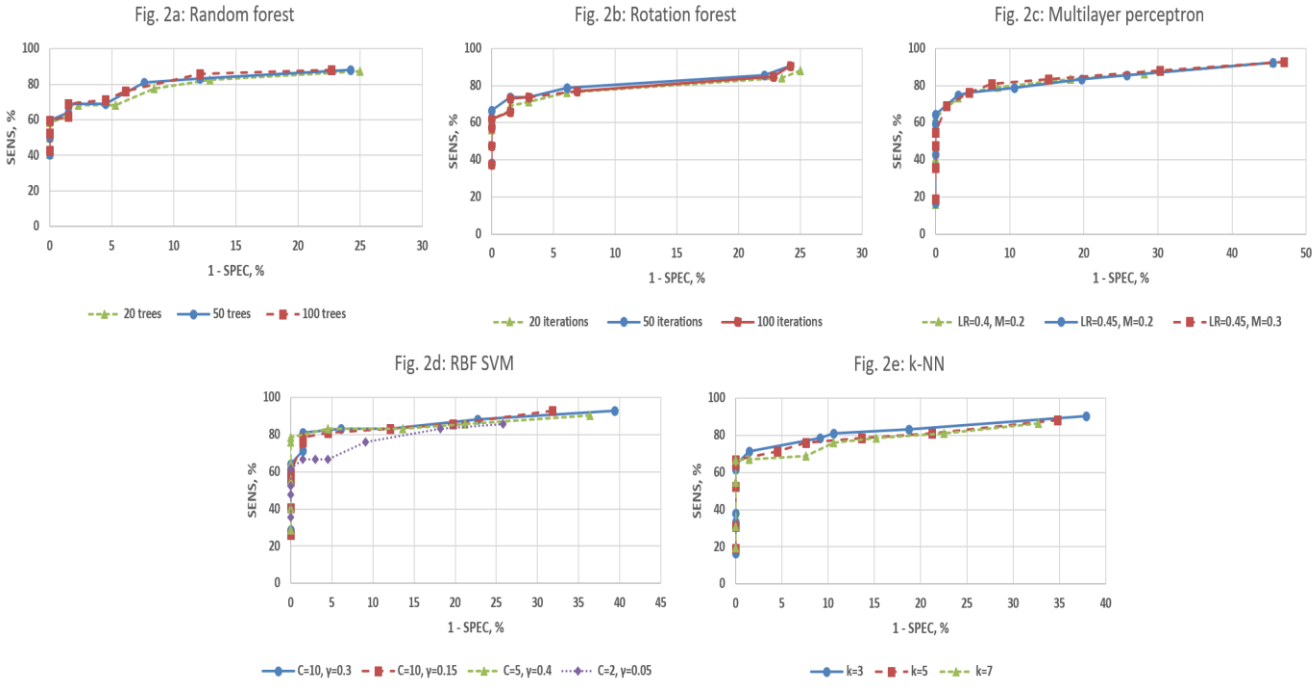

Fig. 2 ROC curves for variations in the classifiers' hyperparameters. Ten points shown for each ROC curve represent different thresholds (from 10% – rightmost point to 100% – leftmost point) for subject-level LOSOCV classification

hyperparameters of the used classifiers. Note that not all the inspected hyperparameter variations are represented in Fig. 2, since this would diminish the readability of the figure. The presented results suggest that the variation in the hyperparameters of a classifier mostly leads to rather small differences in SENS and SPEC, especially in the mid-range of the threshold (around 50%). In the subsequent analyses of the feature selection methods, we used the hyperparameters that achieved the maximum ACC, as given earlier.

Figs. 3 and 4 show the classification results for all classifiers for the symmetrical uncertainty filter, with respect to the number of features included in the dataset, for subject-level and feature vector-level classification, respectively. Generally, a decline in classification accuracy for subject-level classification is observed only when less than 40% of features are retained (except for random forest, where there is no significant decline compared to the starting ACC = 87.0%). For feature vector-level classification, the decline is mostly observed only when less than 30% of features are retained (except for RTF, where a decline is observed when less than 40% are retained). Note that the oscillations in accuracies for LOSOCV subject-level classification are higher than for feature vector-level classification, because the number of instances is smaller (108 subjects vs. 4082 segments). For

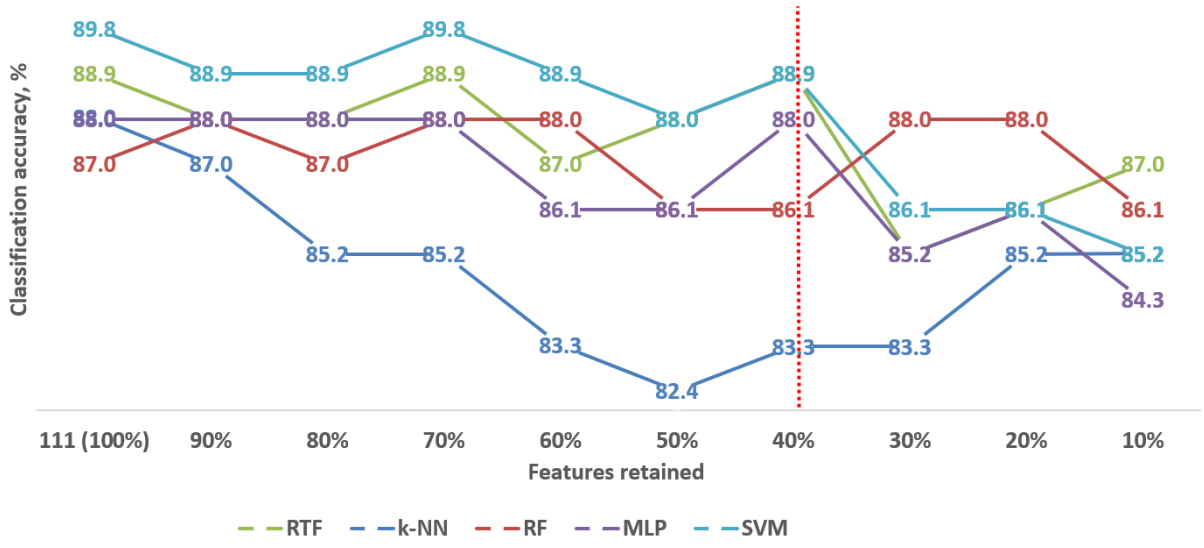

Fig. 3 Classification accuracy (ACC) for the number of retained ranked features selected by symmetrical uncertainty method, subject-level LOSOCV based classification

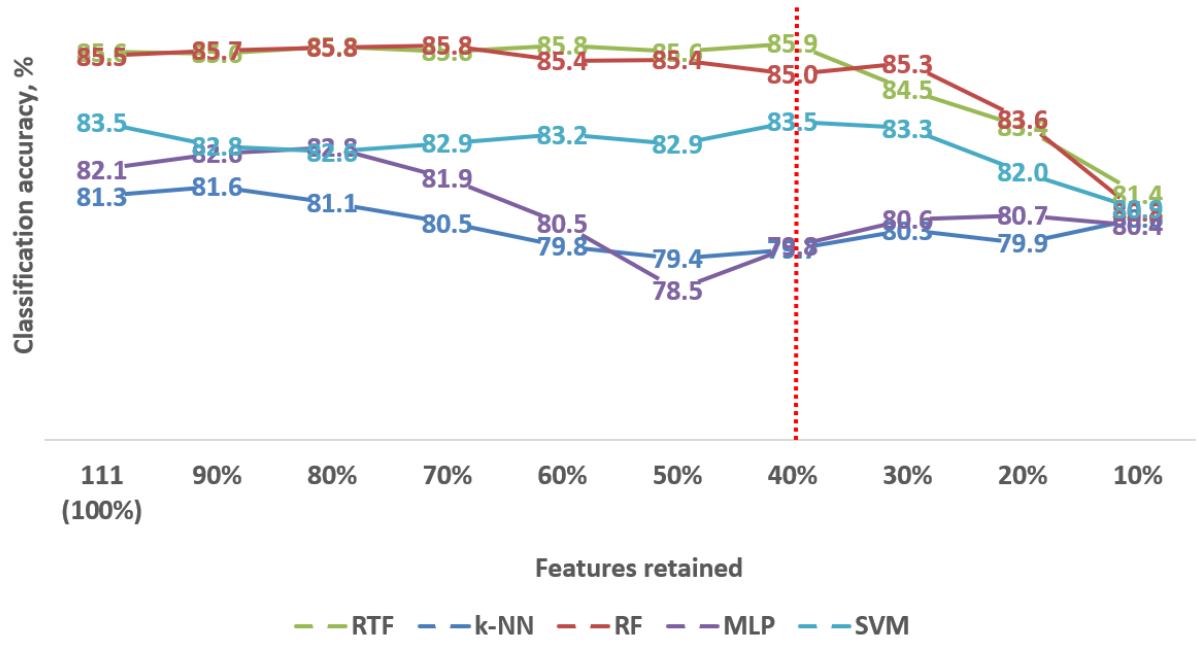

Fig. 4 Classification accuracy (ACC) for the number of retained ranked features selected by symmetrical uncertainty method, feature vector-level LOSOCV based classification

subject-level, misclassification of an individual subject leads to a roughly 1% decline in accuracy. From Figs. 3 and 4, we can conclude that, although all the methods perform relatively similar with respect to absolute accuracy, the results still favor RTF, SVM and RF compared to MLP and *k*-NN. Hence, for comparing the results of Hybrid #1 and Hybrid #2 proposed feature selection methods, we show only the results for the RTF, SVM, and RF classifiers. For conducting the proposed methods, we decided to hold 40% of features (L_{OPT} = 44) as the optimal result for symmetrical uncertainty feature selection, for better comparison purposes among the classifiers (although we could have used a lower threshold for RF).

Figs. 5–7 show comparative subject-level results for the four feature selection methods used in this work for RTF, SVM, and RF, respectively. Fig. 5 shows that, for RTF, Hybrid #2 improves the accuracy and sensitivity of Hybrid #1: ACC (Hybrid #1) = 88.9% vs. ACC (Hybrid #2) = 90.7%, SENS (Hybrid #1) = 73.8% vs. SENS (Hybrid #2) = 78.6%. Overall, Hybrid #2 model is comparable with NB wrapper model, but has only 4 features (18 for NB wrapper).

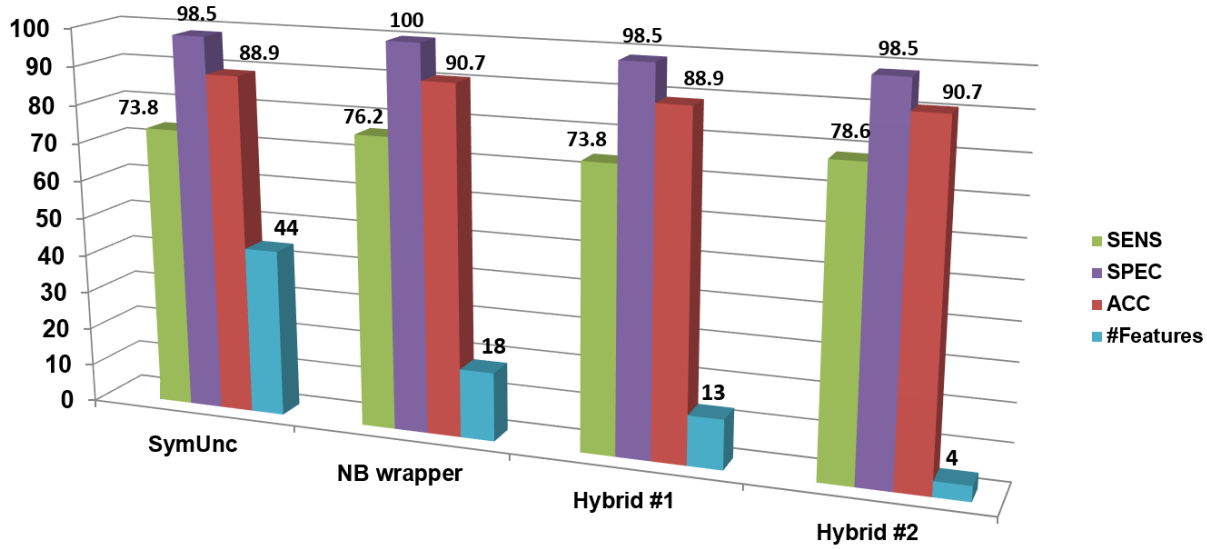

Fig. 5 Rotation forest classification results and the number of retained features (#Features) for the feature selection methods used in this work, LOSOCV subject-level classification

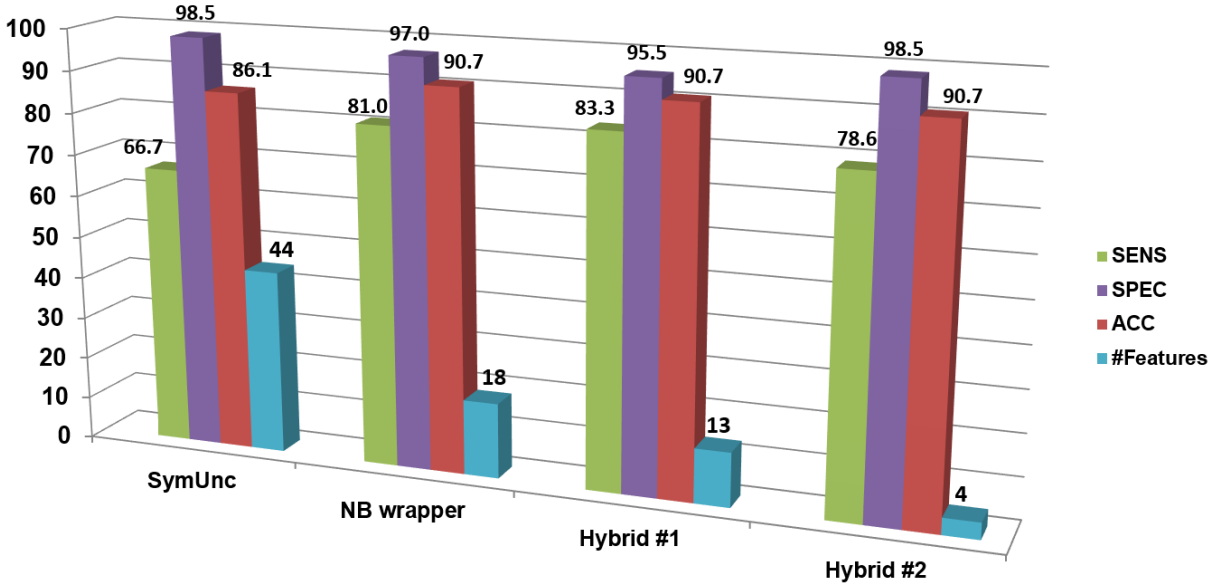

Fig. 6 Support vector machine classification results and the number of retained features (#Features) for the feature selection methods used in this work, LOSOCV subject-level classification

Fig. 6 shows that for SVM, Hybrid #2 retains and even improves the accuracy and sensitivity of Hybrid #1, while having a somewhat smaller specificity: ACC (Hybrid #1) = 87.0% vs. ACC (Hybrid #2) = 88.0%, SENS (Hybrid #1) = 73.8% vs. SENS (Hybrid #2) = 78.6%, SPEC (Hybrid #1) = 95.5%

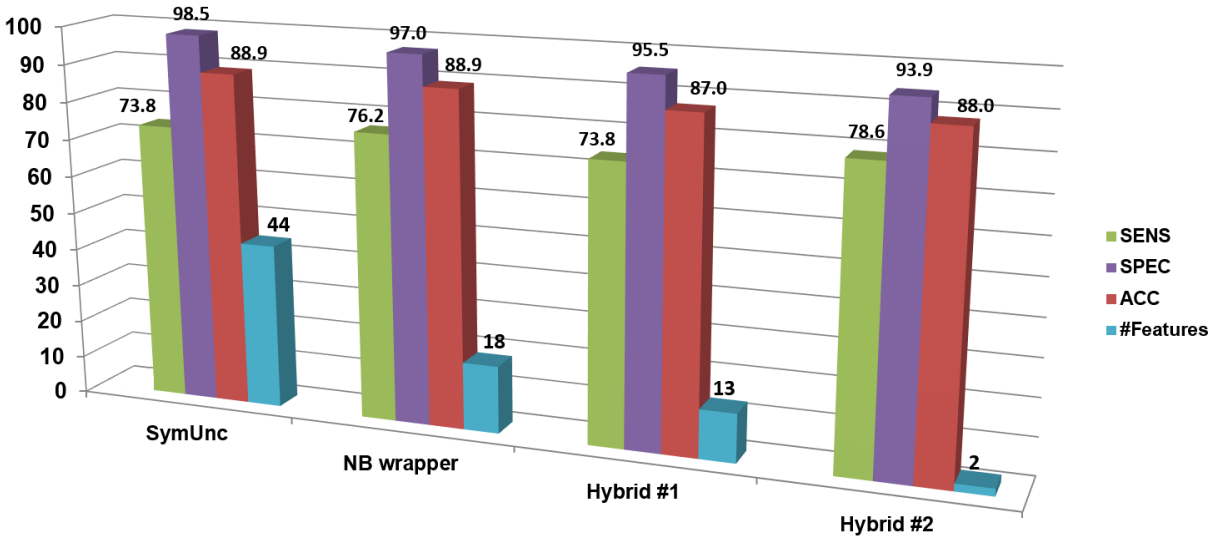

Fig. 7 Random forest classification results and the number of retained features (#Features) for the feature selection methods used in this work, LOSOCV subject-level classification

vs. SPEC (Hybrid #2) = 93.9%. Overall, the results of Hybrid #2 model are comparable with SymUnc and NB wrapper models, but Hybrid #2 has only 2 features (18 for NB wrapper and 44 for SymUnc).

Fig. 7 shows that for RF, Hybrid #2 retains the accuracy, somewhat lowers sensitivity, and somewhat increases specificity of Hybrid #1: ACC (Hybrid #1) = 90.7% vs. ACC (Hybrid #2) = 90.7%, SENS (Hybrid #1) = 83,3% vs. SENS (Hybrid #2) = 78.6%, SPEC (Hybrid #1) = 95.5% vs. SPEC (Hybrid #2) = 98.5%. Overall, the results of Hybrid #2 model are better than SymUnc and comparable to NB wrapper models, but still, Hybrid #2 has only 4 features.

Table 4 summarizes the feature vector-level results for all three classifiers. We can observe that for all three classifiers, the Hybrid #2 results are somewhat reduced compared to the other methods. This is expected, as the number of features is drastically reduced. Note that this reduction is on the order of 1%, while Hybrid #2 method retains only four (or two) features. Table 4 also contains the best achieved results obtained during Hybrid #2 feature selection procedure (recursive feature elimination). All classifiers obtained the best results during Hybrid #2 that were better than most of the results for all the other feature selection methods, while retaining only six or seven features.

Table 4 Feature vector-level LOSOCV classification results for the feature selection methods used in this work, all three classifiers (ACC/SENS/SPEC), %

Classifier	SymUnc	NB wrapper	Hybrid #1	Hybrid #2	Best results during Hybrid #2 / (best no. of features)
RTF	85.9/74.8/92.7	85.8/74.3/93.0	85.4/74.4/92.2	84.7/74.3/91.2	86.3/76.0/92.7 / (6)
SVM	83.5/72.5/90.4	85.4/72.5/93.6	83.8/71.8/91.3	82.1/73.8/87.3	84.0/70.9/92.2 / (7)
RF	85.0/73.5/92.2	85.0/74.6/91.6	85.1/75.1/91.3	83.9/72.6/90.9	85.5/75.6/91.7 / (7)

Figs. 8–10 depict in detail how the classification accuracy behaves when greedy iterative feature removal (Hybrid #2) is performed on the feature set obtained by Hybrid #1, for RTF, SVM, and RF, respectively. The set of 13 features obtained by Hybrid #1 procedure includes (in alphabetical order): AverAlphEn_L , AverAlphEn_Z, AlphEnVar, CVrr, DFA α_2, HaarWavSD_sc8, HF, LF/HF, MaxAlphEn, Rate_B, Rate_E, Rate_U, and TINN.

For RTF, depicted in Fig. 8, note that a significant drop in subject-level classification accuracy occurs only when three features are left in the feature set. Therefore, Hybrid #2 method stops at four selected features. If greedy iterative removal is continued, it eventually leads to the optimal feature for CHF

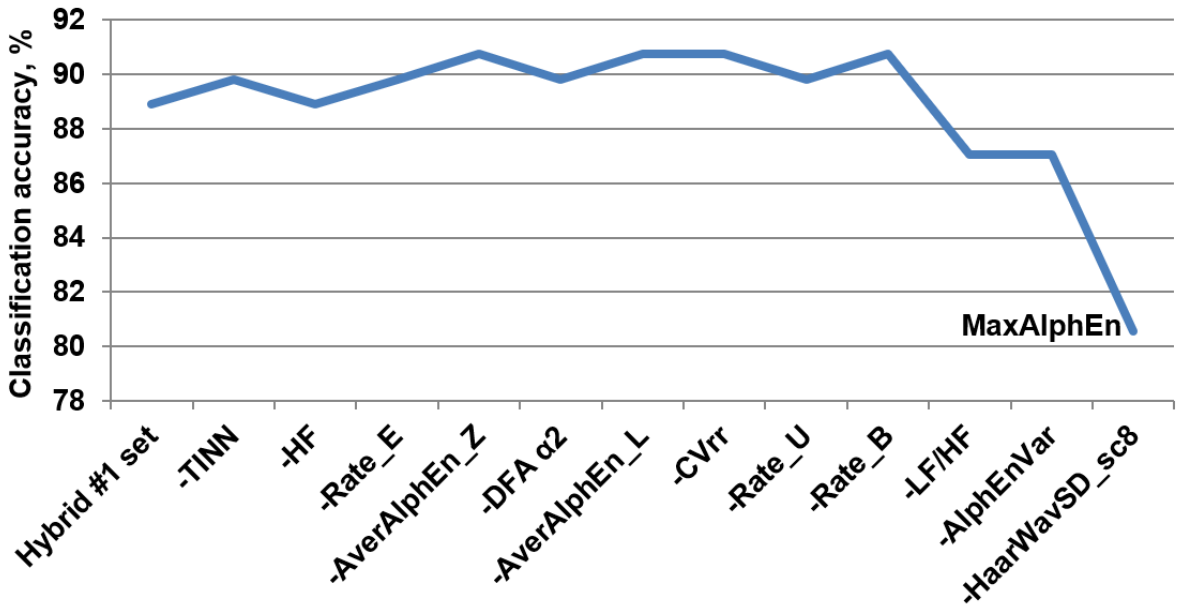

Fig. 8 Detailed analysis of the Hybrid #2's greedy iterative feature removal procedure from Hybrid #1 feature set, subject-level classification for RTF classifier

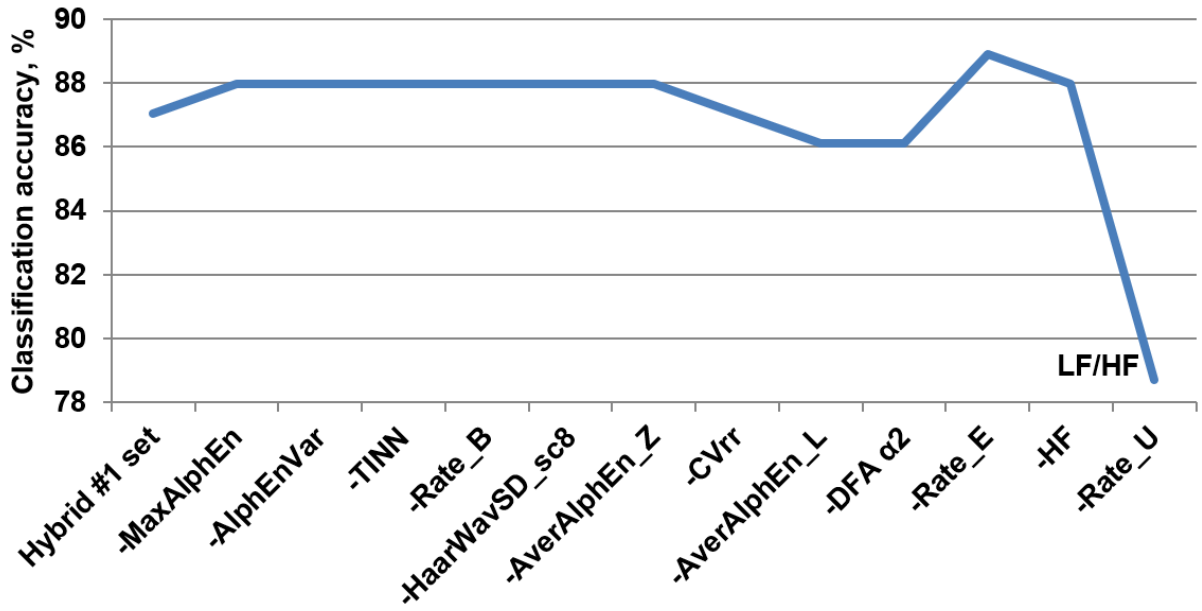

Fig. 9 Detailed analysis of the Hybrid #2's greedy iterative feature removal procedure from Hybrid #1 feature set, subject-level classification for SVM classifier

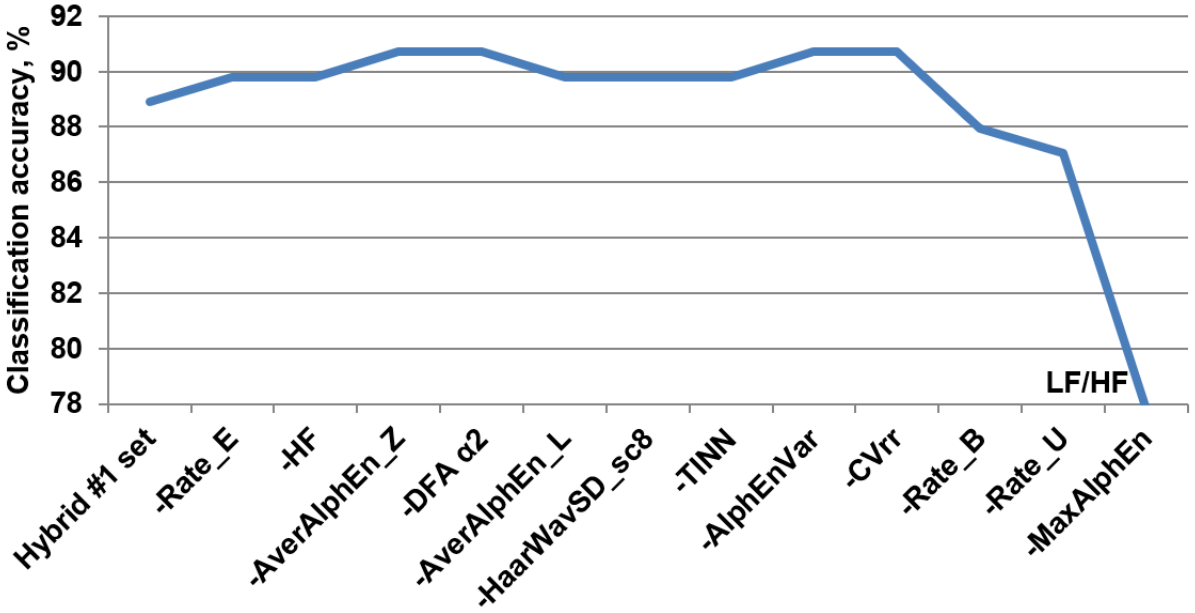

Fig. 10 Detailed analysis of the Hybrid #2's greedy iterative feature removal procedure from Hybrid #1 feature set, subject-level classification for RF classifier

detection in this model, which is MaxAlphEn (maximum alphabet entropy). Using RTF model for only this single feature still results in a very good classification performance (ACC = 80.6%, SENS =

52.4%, SPEC = 98.5%), which means that MaxAlphEn feature alone accurately identifies more than a half of the CHF patients with great certainty when using RTF. For SVM, depicted in Fig. 9, the significant drop in accuracy occurs only at the last feature, which means that Hybrid #2 method stops at two selected featutes: Rate_U and LF/HF. Using SVM for only the single best feature, which is LF/HF, still results in very good classification performance (ACC = 78.7%, SENS = 73.8%, SPEC = 81.8%). For RF, depicted in Fig. 10, the significant drop in accuracy occurs when three features are left in the feature set, which means that Hybrid #2 method stops at four selected features. Greedy iterative removal eventually leads to the optimal feature for automated CHF diagnosis in this model, which is LF/HF. Using RF model for only this single feature results in classification performance (ACC = 77.8%, SENS = 59.5%, SPEC = 89.4%).

The Hybrid #2 model with the best four HRV features in the case of RTF includes: LF/HF, AlphEnVar, MaxAlphEn, and HaarWavSD_sc8. For the SVM, the best model includes Rate_U and LF/HF. Finally, for RF, the best model includes Rate_B, Rate_U, MaxAlphEn, and LF/HF. In Table 5, we present some statistical results for the identified best six features for NSR and CHF groups. It can be observed that all the features show differences between the two groups, yet any single feature is insufficiently accurate to clearly distinguish between the groups.

Table 5 Statistics (first quartile/median /third quartile) for the resulting 6 best features obtained by RTF, SVM and RF methods, compared for NSR and CHF groups

Group	LF/HF	AlphEnVar	MaxAlphEn	HaarWavSD_sc8	Rate_B	Rate_U
NSR	2.221/3.783/6.017	0.157/0.236/0.294	2.493/2.726/2.989	0.074/0.118/0.176	0.034/0.048/0.059	0/0.006/0.018
CHF	0.359/0.834/2.295	0.037/0.124/0.244	2.278/2.611/3.715	0.026/0.046/0.100	0.006/0.023/0.042	0/0/0.002

4 Discussion

In this section, a comparison with related work is presented in order to put the results of this work into current perspective, then some limitations of the current study are outlined, and finally, the obtained model is discussed.

4.1 Comparison with related work

In Table 6, an overview of scientific work related to short-term HRV-based automated discerning of CHF from healthy state is shown. Compared to the other studies, our study included the most records and considered features. The most methodologically similar study is the one by Pecchia et al. [53]. It considers removing the segments with less than 80% of normal beats and employs the same evaluation procedure for subject-level classification (see Section 2.4). However, the study does not report on the location of the analyzed segments in the record nor does it consider editing (interpolating, deleting or filtering) the remaining ectopic beats, which may result in overly optimistic results due to the influence of arrhythmias. Narin et al. [50] consider age and the total of 58 other linear and nonlinear features. By reducing the feature set down to 27 using backward feature elimination, they obtain results that are comparable to our study that uses only four or two features, although they use a somewhat smaller dataset (108 vs. 83). It is important to note than Narin et al. use only the first ectopy-free segment, which may not be representative enough to fully describe a subject's rhythm (we consider segments from two hours and interpolate occasional ectopic beats). The methodology of Isler and Kuntalp [26] includes having demographic information, as well as selecting a single random 5-minute segment from each record, which can lead to unrepresentative results due to omission of a large quantity of important information from subjects. Liu et al. [40] achieved a perfect score on a smaller dataset (47 records in total). However, they do not report on the exact algorithm for removing ectopic beats, interpolation or faulty excerpt removal, which makes the results unreliable. Like the study of Isler and Kuntalp [26], the study of Isler et al. [27] used only a single 5-minute segment for

Table 6 Related work in short-term HRV analysis for discerning CHF from healthy state

Author	Dataset	Features	Classifier	Preprocessing	ACC | SENS | SPEC, %
Pecchia et al. [53]	NSR1 (54 subjects), CHF1 (29 subjects)	Linear (9), nonstandard linear (2) for 24h	CART, (NSR vs. CHF)	5-minute segments, unreported location in record, reported removal of segments if less than 80% of RR intervals in segment is from normal beats; exhaustive search for optimal feature subset	92.8 | 79.3 | 100.0
Narin et al. [50]	NSR1 (54), CHF1 (29)	Age, linear, nonlinear (59 total)	SVM (NSR vs. CHF)	5-minute segments, from beginning of the record first ectopy-free segment, feature backward elimination (reduction from 59 to 27 features)	91.6 | 82.8 | 96.3
Isler and Kuntalp [26]	NSR1 (54), CHF1 (29)	Demographic, linear, Poincare plot, wavelet entropy (30 total)	*k*-NN (NSR vs. CHF)	5-minute segments, arbitrary selection of a 5-min segment from each record, removal of ectopic beats, genetic algorithm search for optimal feature subset (reduction from 30 to 11 features)	96.4 | 100.0 | 94.7
Liu et al. [40]	NSR1 (30), CHF1 (17)	Nonstandard linear and nonlinear (3)	SVM (NSR vs. CHF)	5-minute segments, from beginning of the record, no interpolation nor faulty segment removal reported	100.0 | 100.0 | 100.0
Isler et al. [27]	NSR1 (54), CHF1 (29)	Linear (41), nonlinear (17)	Three-stage: perceptron first two stages, complex classifier in the final stage, RBF ANN best model	5-minute segments, only the first 5-minute segment from each record, deletion used for removing ectopic beats, feature selection based on t-test at 5% threshold (reduction from 58 to 34 features)	98.8 | 100.0 | 98.1
This study	NSR1 (49), NSR2 (17), CHF1 (27), CHF2 (15)	Linear, nonlinear (111 total)	RTF, RF, SVM, (NSR vs. CHF)	5-minute segments, 2h from the beginning of the records, removal of segments with less than 90% of normal beats, linear interpolation of ectopic beats, hybrid feature selection	90.7 | 78.6 | 98.6 (RF and RTF)

classification between healthy subjects and CHF patients. A multi-stage classification scheme, consisting of simple perceptrons in the first two stages and a more complex classifier in the final stage achieved high accuracy of 98.8%. While the results seem promising, we consider that the following problems may have led to overly optimistic results: 1) selecting only a single segment from a record (this bias was avoided by Pecchia et al. [53] and in our study using LOSOCV classification), 2) using

deletion for removing ectopic beats, which is not the best way to handle ectopic beats [53], and 3) using a smaller set of records.

It can be concluded that the main concern with all the studies are the data preprocessing methods. Most studies do not report on editing ectopic beats, which is not in accordance with medical guidelines [70] and other expert recommendations [54], and which may lead to optimistic results that may not be achievable in practice. Since CHF segments usually contain more ectopic beats than the segments from NSR subjects, it is much more likely to achieve perfect classification that is questionable in practice due to the existence of arrhythmias. We hold that our approach leads to a generalizable and efficient model for CHF diagnosis from short-term HRV that is more realistic in practice and that may be used to identify with certainty (high SPEC) many, but not all (moderately high SENS) CHF patients. We also note that our study is currently the only one that uses all four available NSR and CHF databases for short-term HRV analysis, and thus provides more accurate quantification of the problem due to the larger sample size. The influence of the two additional included databases on the results is significant, especially since CHF2 database contains records from severe heart failure patients, which were omitted by other studies.

For comparison purposes, we also provide a selection of the best currently available studies for long-term HRV-based discerning of CHF from healthy state in Table 7. The inspected studies mostly analyzed 24h of recordings, with some exceptions (e.g. Yu and Lee [79] analyzed only the first 4096 s of recordings). It should be noted that it is expected that long-term HRV-based automated detection of CHF can achieve much better classification results than the short-term detection, because long-term measurements include more information about the heart rhythm changes present in the circadian cycle, which is known to be affected in CHF [1, 34, 65]. From methodological standpoint, most long-term HRV studies seem to be better conducted than the short-term studies. Most of the studies included all four available NSR and CHF databases from PhysioNet, but still, most studies do not report any editing of ectopic beats (except for Yu and Lee [79]), while some studies (e.g. Chen et al. [7]) also do not report on removing the segments that contain many ectopic beats. Both deficiencies may lead to

Table 7 Studies in long-term HRV analysis for discerning CHF from healthy state

Author	Dataset	Features	Classifier	Preprocessing	ACC | SENS | SPEC, %
Asyali [4]	NSR1 (54 subjects), CHF1 (29 subjects)	Linear time and frequency (9), Welch periodogram	Linear discriminant analysis (NSR vs. CHF)	24h segments, removal of records if less than 90% of RR intervals are from normal beats, no interpolation of ectopic beats reported, cubic interpolation for PSD Welch periodogram calculation	93.2 | 81.8 | 98.1
Mellilo et al. [45]	NSR1 (53), NSR2 (18), CHF1 (26), CHF2 (13)	Linear time and frequency domain (13), Welch and Lomb Scargle periodogram	CART (NSR vs. CHF)	24h segments, removal of records if less than 85% of RR intervals are from normal beats, no interpolation of ectopic beats reported, cubic interpolation for PSD Welch periodogram calculation, exhaustive search feature selection	96.4 | 89.7 | 100.0
Mellilo et al. [46]	CHF1(11), CHF1 and CHF2(30)	Linear time and frequency domain (13), Lomb Scargle periodogram	CART (mild CHF vs severe CHF)	24h segments, removal of records if less than 80% of RR intervals are from normal beats, no interpolation of ectopic beats reported, exhaustive search feature selection	85.4 | 93.3 | 63.6
Yu and Lee [79]	NSR1 (54), CHF1 (29)	Age and gender, linear time and frequency (16), bi-spectrum (32), total 50 features	RBF SVM (NSR vs. CHF)	4096 s from beginning of the record, removal of ectopic beats and trends, genetic algorithm search for optimal feature subset (reduction from 50 to 11–14 features, depending on the trial), no faulty segment removal reported	98.8 | 96.6| 100.0
Shahbazi and Asl [67]	CHF1 (10), CHF2 (29)	Linear time and frequency (14), FFT spectrum estimate, and nonlinear (11)	*k*-NN (*k*=5) (mild CHF vs severe CHF)	24h segments, removal of records if less than 80% of RR intervals are from normal beats, no interpolation of ectopic beats reported, GDA dimensionality reduction.	100.0 | 100.0 | 100.0
Chen et al. [7]	NSR1 (54), NSR2 (18), CHF1 (29) CHF2 (15)	Total of 180 static and dynamic linear time and frequency, nonlinear features	DT-SVM, 4 classes: NSR, mild CHF, moderate CHF, severe CHF	24h segments with 5-minute segments hybrid analysis, no interpolation nor faulty segment removal reported, RR intervals longer than 3 s discarded, backward elimination feature selection (reduction from 180 to 23 features)	96.6 | N/A | N/A
Mahajan et al. [41]	69 NSR (NSR1 and NSR2), 38 CHF (CHF1 and CHF2)	Probabilistic symbolic pattern recognition (8 features), long scale AVNN, SDNN	Random forest	24h segments, removal of records if less than 90% of RR intervals are from normal beats, no interpolation of ectopic beats reported, no feature selection	98.1 | 94.7 | 100.0

to somewhat better results than that can be achieved in practice.

Regarding the use of feature selection methods, we would like to point out that other researchers (as seen in Tables 6 and 7) used exhaustive search when the number of features was small (e.g. Pecchia et al. [53], Melillo et el. [45, 46]), genetic algorithm based wrapper approach (e.g. Isler and Kuntalp [26]: reduction from 30 to 11 features, Yu and Lee [79]: from 50 to 11−14 features, depending on the trial) and backward feature elimination wrapper (e.g. Narin et al. [50]: reduction from 59 to 27 features, Chen et al. [7]: reduction from 180 to 23 features). While most of the feature selection methods they used appear to be simpler than our approach, still, the resulting feature set is not small enough to allow adequate interpretation of the model, especially when exhaustive feature selection is not used. Hence, the advantage of our methodology compared to the other works also lies in obtaining only a small number of highly informative features that allow for physiological interpretation of the obtained models (see section 4.3).

4.2 Limitations

The study was conducted on a relatively small sample of CHF patients (42), although still larger than all the other studies shown in Table 6. The included sample is the maximum number of freely available patient records having CHF from the PhysioNet portal [16]. While there have been several studies and clinical trials that included more CHF patients (200 or more), all of them conducted with high quality 24h Holter ECG monitoring, the valuable datasets are currently not publicly available [19, 38, 43], which prevents independent evaluations. Also, the limitation of the freely available datasets is the sampling rate of 128 Hz or 250 Hz, which constrains the temporal resolution of true R peak occurrence to within roughly ±4 ms or ±2 ms, respectively, from the detected one. While this was shown to be acceptable for normal subjects [8], it may introduce a bias in spectrum and in some nonlinear features, more so for patients with severe CHF, where HRV is usually diminished [8, 47].

While the quantification of the bias is difficult, we note that LF/HF spectral feature was found to be significant in our study by all the used classifiers, regardless of the sampling rate limitations.

4.3 Model interpretation and feature set

The Hybrid #2 model for RTF includes four HRV features: LF/HF, AlphEnVar, MaxAlphEn, and HaarWavSD_sc8, SVM includes two HRV features: Rate_U and LF/HF, while RF includes four features: Rate_B, Rate_U, MaxAlphEn, and LF/HF. The significance of LF/HF ratio, established by all three classifiers, is unsurprising, as the power in the LF band can be influenced by vagal, sympathetic, and baroreflex mechanisms depending on the context, whereas HF power is produced by the efferent vagal activity due to respiratory activity [34, 66]. Vagal activity at least is clearly affected in CHF [14], thus enabling its discerning from NSR. Alphabet entropy features AlphEnVar and MaxAlphEn both reflect the way in which the time series changes in very short-term period, consisting of only four sequential RR intervals [33]. A selected threshold of alphabet entropy's no-change of only 20 ms enables high sensitivity in measuring the increase or decrease in RR interval durations, while eliminating the problem of the low sampling frequencies. AlphEnVar is the variance of alphabet entropy series of the original series, hence showing that lower dispersion in sequential RR-interval changes is indicative of CHF (see Table 5). Sequential RR-interval changes are reduced in CHF (interval changes are usually within the 20 ms threshold), because CHF affected heart has more difficulty in adapting to the usual environmental changes than the healthy one. While reduced HRV complexity in CHF patients is well-established in some traditional HRV features, for example SDNN [49], here we have also shown that a very-short term complexity measures (i.e. AlphEnVar) also demonstrates the heart’s reduced dynamics in CHF, significantly differing from the healthy state. MaxAlphEn, on the other hand, measures the largest alphabet entropy of four sequential RR measurements among all alphabet entropies in a 5-minute segment. Here also, a higher maximum entropy indicates NSR (see Table 5), which is also in accordance with the observation that healthy heart shows larger dynamical complexity. Rate_B and Rate_U are also both alphabet entropy method

related features, but unlike AlphEnVar and MaxAlphEn, they are qualitative features in the sense that they measure the rate of occurrence of rhythm changes in a segment, which is determined using three levels: shortening (-), prolongation (+), and no-change of an RR interval [33]. The letter "B" in Rate_B represents the pattern "0 0 +" occurring in a segment, while letter "U" in Rate_U represents the pattern "+ + -". As we have previously shown in the study [33], letter "B" is typically encountered in NSR. Also, as we show here, letter "B" is more common for NSR than for CHF, indicating that RR-interval prolongation after no-change in rhythm is more common in healthy subjects compared to CHF patients. Normal RR-interval prolongation is usually linked to normal respiratory rhythm, which may be diminished in CHF, especially in severe heart failure. However, letter "U" was previously not shown to be typical for NSR or for any studied cardiac arrhythmia. As it can be seen from Table 5, letter "U" occurrence is still much more common in NSR than in CHF. The reason may be because the heart of a NSR subject can adjust more quickly to environmental changes than the heart of a CHF patient. Letter "U" pattern may also be present in respiratory sinus arrhythmia, which is clinically non-significant if related to respiratory rhythm and is also more common in healthy subjects. Standard deviation of Haar Wavelet of RR-interval series is a generalization of the RMSSD measurement [71]. HaarWavSD_sc8 is measured at scale 8. The feature was already shown to be efficient in discriminating healthy subjects from CHF patients, albeit for long time series [71]. In this work, we also demonstrate that HaarWavSD_sc8 is generally lower for CHF group in short-term setting, which is also consistent with reduced dynamical complexity in CHF patients.

All three classifiers use a black-box (or a complex function in the case of SVM) classification model. Hence, the exact ranges for the resulting features that constitute the optimal classification model cannot be visualized nor presented in the form of clear rules other than the statistics shown in Table 5. Clear rules can be achieved using individual C4.5 tree, but accuracy usually drops in this case compared to forests. However, the goal of the study was to test the accuracy limits of models for automatic detection of CHF from short-term HRV as well as to provide clear interpretation of the features used in the best models.

We note that some of the features given in Table 2 may be considered by some to be unwarranted for the analysis (e.g. multiscale sample entropy at larger scales or VLF, due to potentially insufficient number of RR intervals for an accurate estimate). However, as the goal of this work was to use machine learning methods to filter out irrelevant or redundant features, such features would be eliminated during the analysis. We contend that it cannot be expected from researchers to manually construct a perfect combination of features for a given problem, so at least some redundancy in the feature set can be expected. As we show in this work, a quite accurate classification may be achieved by using 40% of the entire feature set using a simple feature filtering procedure (i.e. symmetrical uncertainty), while more elaborate procedures (Hybrid #2) lead to an accurate combination with a drastic reduction to only four (or two) features (roughly 3.5-1.7% of the original feature set size).

5 Conclusion

An efficient hybrid feature selection algorithm (Hybrid #2) was proposed for automated CHF detection from short-term HRV segments. The method succeeded in either improving (for RTF and RF) or retaining (for SVM) the initial feature set accuracy as well as greatly reducing the feature set. The choice of the best features and the classification results are similar between the three classifiers, so the use of any of the three classifiers can be recommended. The obtained results suggest that this method may be used for first-hand screening for CHF. The method may complement the additional, more medically demanding tests (i.e. echocardiography, natriuretic peptides). Additionally, this approach may be beneficial for other similar medical problems, particularly in diagnosis or automated detection scenarios in biomedicine. The resulting model also confirms the usefulness of recently proposed alphabet entropy features in HRV analysis [33].

Future studies should focus on evaluating the obtained model on an even larger and higher quality dataset, when such a one becomes publicly available. Also, we plan to investigate the design of a medical expert system based on the obtained model that would support remote and portable diagnosis

of CHF, possibly via handheld devices [50]. Additionally, motivated by the related studies, we plan to investigate some features that may also be used for short-term HRV analysis (e.g. bispectral features [79], different wavelet functions entropies [26]), in hope of achieving even a slightly higher accuracy rate using the proposed methodology. Finally, since the currently proposed complex hybrid feature selection method was successful in obtaining very small and highly informative feature sets, future work may include comparison of the approach with other feature selection approaches on various biomedical datasets.

Figures: The Authors prefer that color is not used for reproduction of figures in print.

Conflict of Interest: None declared.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Acharya RU, Joseph KP, Kannathal N, Lim CM, Suri JS (2006) Heart Rate Variability: a Review. Med Biol Eng Comput 44(12):1031–1051. DOI: 10.1007/s11517-006-0119-0

[2] Aktaruzzaman M, Migliorini M, Tenhunen M, Himanen SL, Bianchi AM, Sassi R (2015) The addition of entropy-based regularity parameters improves sleep stage classification based on heart rate variability. Med Biol Eng Comput 53(5):415–425. DOI: 10.1007/s11517-015-1249-z

[3] Asl BM, Setarehdan SK, Mohebbi M (2008) Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. Artif Intell Med 44(1):51–64. DOI: 10.1016/j.artmed.2008.04.007

[4] Aslayli MH (2003) Discrimination Power of Long-Term Heart Rate Variability Measures. In: Proc 25th Int Conf of the IEEE EMBS, pp. 200–203. DOI: 10.1109/IEMBS.2003.1279568

[5] Bauer A, Kantelhardt JW, Barthel P, Schneider R, Mäkikallio T, Ulm K, Hnatkova K, Schömig A, Huikuri HV, Bunde A, Malik M, Schmidt G (2006) Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. Lancet 367(9523):1674–1681. DOI: 10.1016/S0140-6736(06)68735-7

[6] Breiman L (2001) Random forests. Mach Learn 45:5–32.

[7] Chen W, Zheng L, Li K, Wang Q, Liu G, Jiang Q (2016) A Novel and Effective Method for Congestive Heart Failure Detection and Quantification Using Dynamic Heart Rate Variability Measurement. PLOS ONE 11(11): e0165304. DOI: 10.1371/journal.pone.0165304

[8] Clifford GD, Tarassenko L (2005) Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. IEEE Trans Biomed Eng 52(4):630–638. DOI: 10.1109/TBME.2005.844028

[9] Costa M, Goldberger AL, Peng C-K (2005) Broken Asymmetry of the Human Heartbeat: Loss of Time Irreversibility in Aging and Disease. Phys Rev Lett 95:198102. DOI: 10.1103/PhysRevLett.95.198102

[10] Dos Santos L, Barroso JJ, Macau EEN, De Godoy MF (2015) Assessment of heart rate variability by application of central tendency measure. Med Biol Eng Comput 53(11):1231–1237. DOI: 10.1007/s11517-015-1390-8

[11] Faust O, Prasad VR, Swapna G, Chattopadhyay S, Lim TC (2012) Comprehensive analysis of normal and diabetic heart rate signals: a review. J Mech Med Biol 12:1240033. DOI: 10.1142/S0219519412400337

[12] Faust O, Acharya RU, Krishnan SM, Min LC (2004) Analysis of Cardiac Signals Using Spatial Filling Index and Time-Frequency Domain. Biomed Eng Online 3:30. DOI: 10.1186/1475-925X-3-30

[13] Fernandez-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we Need Hundreds of Classiers to Solve Real World Classication Problems? J Mach Learn Res 15:3133–3181.

[14] Floras JS (2009) Sympathetic Nervous System Activation in Human Heart Failure. J Am Coll Cardiol 54(5):375–385. DOI: 10.1016/j.jacc.2009.03.061

[15] Friedman N, Geiger D, Goldszmidt M (1997) Bayesian Network Classifiers. Mach Learn 29(2–3) 131–163. DOI: 10.1023/A:1007465528199

[16] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215–e220. DOI: 10.1161/01.CIR.101.23.e215

[17] Goode KM, Clark AL, Cleland JGF (2008) Ruling out heart failure in primary-care: The cost-benefit of pre-screening using NT-proBNP and QRS width. Int J Cardiol 130:426–437. DOI: 10.1016/j.ijcard.2007.08.131

[18] Guidi G, Pollonini L, Dacso CC, Iadanza E (2015) A multi-layer monitoring system for clinical management of Congestive Heart Failure. BMC Med Inform Decis 15(Suppl 3):S5. DOI: 10.1186/1472-6947-15-S3-S5

[19] Guzzetti S, La Rovere MT, Pinna GD, Maestri R, Borroni E, Porta A, Mortara A, Malliani A (2005) Different spectral components of 24 h heart rate variability are related to different modes of death in chronic heart failure. Eur Heart J 26:357–362. DOI: 10.1093/eurheartj/ehi067

[20] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA Data Mining Software: an Update. SIGKDD Explor Newsl 11(1):10–18. DOI: 10.1145/1656274.1656278

[21] Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall: Upper Saddle River, NJ, USA.

[22] Ho YL, Lin C, Lin YH, Lo MT (2011) The prognostic value of non-linear analysis of heart rate variability in patients with congestive heart failure—a pilot study of multiscale entropy. PLoS One 6(4):e18699. DOI: 10.1371/journal.pone.0018699

[23] Hoque N, Bhattacharyya DK, Kalita JK (2014) MIFS-ND: A mutual information-based feature selection method. Expert Syst Appl 41(14):6371–6385. DOI: 10.1016/j.eswa.2014.04.019

[24] Huikuri HV, Perkiömäki JS, Maestri R, Pinna GD (2009) Clinical impact of evaluation of cardiovascular control by novel methods of heart rate dynamics. Phil Trans R Soc A 367:1223–1238. DOI: 10.1098/rsta.2008.0294

[25] Hutchinson TP (2003) Statistics and graphs for heart rate variability: pNN50 or pNN20? Physiol Meas 24(3):N9–14.

[26] Isler Y, Kuntalp M (2007) Combining Classical HRV Indices with Wavelet Entropy Measures Improves to Performance in Diagnosing Congestive Heart Failure. Comput Biol Med 37(10):1502–1510. DOI: 10.1016/j.compbiomed.2007.01.012

[27] Isler, Y, Narin A, Ozer M, Perc M. (2019) Multi-stage classification of congestive heart failure based on short-term heart rate variability. Chaos, Solitons and Fractals 118:145–151.DOI: 10.1016/j.chaos.2018.11.020

[28] Jovic A, Bogunovic N (2011) Electrocardiogram Analysis Using a Combination of Statistical, Geometric, and Nonlinear Heart Rate Variability Features. Artif Intell Med 51(3):175–186. DOI: 10.1016/j.artmed.2010.09.005

[29] Jovic A, Bogunovic N (2012) Evaluating and Comparing Performance of Feature Combinations of Heart Rate Variability Measures for Cardiac Rhythm Classification. Biomed Signal Process Control 7(3):245–255. DOI: 10.1016/j.bspc.2011.10.001

[30] Jovic A, Bogunovic N, Cupic M. (2013) Extension and Detailed Overview of the HRVFrame Framework for Heart Rate Variability Analysis. In: Proc IEEE EuroCon 2013 Conf, pp. 1757–1763. DOI: 10.1109/EUROCON.2013.6625215

[31] Jovic A, Brkic K, Bogunovic N (2015) A review of feature selection methods with applications. In: Proc Int Conf MIPRO 2015, pp. 1447–1452. DOI: 10.1109/MIPRO.2015.7160458

[32] Jovic A, Bogunovic N (2017) HRVFrame framework. Available at: http://www.zemris.fer.hr/~ajovic/hrvframe/hrvframe.html. Accessed: 11 Feb 2017.

[33] Jovic A, Jovic F (2017) Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series. Biomed Signal Process Control 31:217–230. DOI: 10.1016/j.bspc.2016.08.010

[34] Krstacic G, Gamberger D, Krstacic A, Smuc T, Milicic D (2008) The Chaos Theory and Non-linear Dynamics in Heart Rate Variability in Patients with Heart Failure. In: 2008 Computer in Cardiology Conference, pp. 957–959. DOI: 10.1109/CIC.2008.4749202

[35] Krstacic G, Krstacic A, Smalcelj A, Milicic D, Jembrek-Gostovic M (2007) The «Chaos Theory» and Non-linear Dynamics in Heart Rate Variability Analysis: Does it Work in Short Time Series in Patients with Coronary Heart Disease? Ann Noninvasive Electrocardiol 12(2):130–136. DOI: 10.1111/j.1542-474X.2007.00151.x

[36] Krstacic G, Parati G, Gamberger D, Castiglioni P, Krstacic A, Steiner R (2012) Heart rate variability and nonlinear dynamic analysis in patients with stress-induced cardiomyopathy. Med Biol Eng Comput 50:1037–1046. DOI: 10.1007/s11517-012-0947-z

[37] Kuncheva LI, Rodriguez JJ (2007) An Experimental Study on Rotation Forest Ensembles. In: Proc 7th Int Conf on Multiple Classifier Systems MCS'07, pp. 459–468. DOI: 10.1007/978-3-540-72523-7_46

[38] La Rovere MT, Pinna GD, Maestri R, et al. for the GISSI-HF Investigators (2012) Autonomic markers and cardiovascular and arrhythmic events in heart failure patients: still a place in prognostication? Data from the GISSI-HF trial. Eur J Heart Fail 14:1410–1419. DOI: 10.1093/eurjhf/hfs126

[39] Liu H, Setiono R (1996) A probabilistic approach to feature selection - A filter solution. In: Proc 13th Int Conf Mach Learn ICML 1996, pp. 319–327.

[40] Liu G, Wang L, Wang Q, Zhou G, Wang Y, Jiang Q (2014) A New Approach to Detect Congestive Heart Failure Using Short-Term Heart Rate Variability Measures. PLoS One 9(4):e93399. DOI: 10.1371/journal.pone.0093399

[41] Mahajan R, Viangteeravat T, Akbilgic O (2017) Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics. Int J Med Inform 108:55–63, 2017. DOI: 10.1016/j.ijmedinf.2017.09.006

[42] Maestri R, Pinna GD, Accardo A, et al. (2007) Nonlinear indices of heart rate variability in chronic heart failure patients: redundancy and comparative clinical value. J Cardiovasc Electrophysiol 18:425–433. DOI: 10.1111/j.1540-8167.2007.00728.x

[43] Mäkikällio TH, Huikuri HV, Hintze U, et al. for the DIAMOND Study Group (2001) Fractal analysis and time- and frequency-domain measures of heart rate variability as predictors of mortality in patients with heart failure. Am J Cardiol 87:178–182. DOI: 10.1016/S0002-9149(00)01312-6

[44] Masetic Z, Subasi A (2016) Congestive heart failure detection using random forest classifier, Comput Methods Programs Biomed 130:54–64. DOI: 10.1016/j.cmpb.2016.03.020

[45] Melillo P, Fusco R, Sansone M, Bracale M, Pecchia L (2011) Discrimination power of long-term heart rate variability measures for chronic heart failure detection. Med Biol Eng Comput 49(1):67–74. DOI: 10.1007/s11517-010-0728-5

[46] Melillo P, De Luca N, Bracale M, Pecchia L (2013) Classification Tree for Risk Assessment in Patients Suffering from Congestive Heart Failure via Long-Term Heart Rate Variability. IEEE J-BHI 17(3):727–733.

[47] Merri M, Farden DC, Mottley JG, Titlebaum EL (1990) Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. IEEE Trans Biomed Eng 37:99–106. DOI: 10.1109/10.43621

[48] Mitchell TM (1997) Machine Learning. McGraw-Hill: New York, USA.

[49] Musialik-Łydka A, Sredniawa B, Pasyk S (2003) Heart rate variability in heart failure. Kardiol Pol 58(1):10-16.

[50] Narin A, Isler Y, Ozer M (2014) Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance. Comput Biol Med 45:72–79. DOI: 10.1016/j.compbiomed.2013.11.016

[51] Nezu A, Kimura S, Kobayashi T, Osaka H, Uehara S (1996) Coefficient of variation of R-R intervals in severe brain damage. Brain Dev 18(6):453–455.

[52] Park J, Kang K (2014) PcHD: Personalized classification of heart beat types using a decision tree. Comput Biol Med 54:79–88. DOI: 10.1016/j.compbiomed.2014.08.013

[53] Pecchia L, Melillo P, Sansone M, Bracale M (2011) Discrimination Power of Short Term Heart Rate Variability Measures for CHF Assessment. IEEE Trans Inf Technol Biomed 15(1):40–46. DOI: 10.1109/TITB.2010.2091647

[54] Peltola MA (2012) Role of editing of R–R intervals in the analysis of heart rate variability. Front Physiol 3:148. DOI: 10.3389/fphys.2012.00148

[55] Peng RC, Zhou X-L, Lin W-H, Zhang Y-T (2015) Extraction of Heart Rate Variability from Smartphone Photoplethysmograms. Comput Math Methods Med 2015:516826. DOI: 10.1155/2015/516826

[56] Pérez NP, Guevara López MA, Silva A, Ramos I (2015) Improving the Mann-Whitney statistical test for feature selection: an approach in breast cancer diagnosis on mammography. Artif Intell Med 63(1):19–31. DOI: 10.1016/j.artmed.2014.12.004

[57] Perkiömäki JS, Zareba W, KalariaVG, Couderc J, Huikuri HV, Moss AJ (2001) Comparability of nonlinear measures of heart rate variability between long- and short-term electrocardiographic recordings. Am J Cardiol 87:905–908. DOI: 10.1016/S0002-9149(00)01537-X

[58] Perkiömäki JS, Thomsen BPE, Kiviniemi AM, Messier MD, Huikuri HV (2011) Risk factors of self-terminating and perpetuating ventricular tachyarrhythmias in post-infarction patients with moderately depressed left ventricular function, a CARISMA subanalysis. Europace 13:1604–1611. DOI: 10.1093/europace/eur166

[59] Pinnacchio G, Lanza GA, Stazi A, Careri G, Coviello I, Mollo R, Crea F (2015) Determinants of heart rate turbulence in individuals without apparent heart disease and in patients with stable coronary artery disease. Europace 17(12):1855–1861. DOI: 10.1093/europace/euu338

[60] Platt J (1998) Fast training of support vector machines using sequential minimal optimization. In: Advances in kernel methods–support vector learning. MIT Press: Cambridge, MA, USA, pp. 185–208.

[61] Pocock SJ, Ariti CA, McMurray JJV, 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. DOI: 10.1093/eurheartj/ehs337

[62] Porta A, Baselli G, Liberati D, Montano N, Cogliati C, Gnecchi-Ruscone T, Cerutti S (1998) Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. Biol Cybern 78(1):71–78. DOI: 10.1007/s004220050414

[63] Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 278:H2039–2049.

[64] Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation forest: A new classifier ensemble method. IEEE Trans Pattern Anal Mach Intell 28(10)1619–1630. 10.1109/TPAMI.2006.211

[65] Sassi R, Cerutti S, Lombardi F, Malik M, Huikuri HV, Peng C-K, Schmidt G, Yamamoto Y (2015) Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. Europace 17:1341–1353. DOI: 10.1093/europace/euv015

[66] Shaffer F, McCraty R, Zerr CL (2014) A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability, Front. Psychol. 5:340. DOI: 10.3389/fpsyg.2014.01040

[67] Shahbazi F, Asl BM (2015) Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability. Comput Methods Programs Biomed 122(2):191–198. DOI: 10.1016/j.cmpb.2015.08.007

[68] Shin JH, Hwang SH, Chang MH, Park KS (2011) Heart rate variability analysis using a ballistocardiogram during Valsalva manoeuvre and post exercise. Physiol Meas 32(8):1239–1264. DOI: 10.1088/0967-3334/32/8/015

[69] Singh JP, Larson MG, O'Donell CJ, Wilson PF, Tsuji H, Lyod-Jones DM, Levy D (2000) Association of hyperglycemia with reduced heart rate variability: the Framingham heart study. Am J Cardiol 86:309–312.

[70] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology (1996) Heart Rate Variability Guidelines: Standards of Measurement, Physiological Interpretation, and Clinical Use. Eur Heart J 17(3):354–381.

[71] Teich MC (1998) Multiresolution Wavelet Analysis of Heart Rate Variability for Heart-Failure and Heart-Transplant Patients. In: Proc 20th Ann Int Conf IEEE/EMBS, pp. 1136–1141. DOI: 10.1109/IEMBS.1998.747071

[72] The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) (2016) 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 37:2129–2200. DOI: 10.1093/eurheartj/ehw128

[73] Treo EF, Herrera MC, Valentinuzzi ME (2005) Algorithm for identifying and separating beats from arterial pulse records. Biomed Eng Online 4:48. DOI: 10.1186/1475-925X-4-48

[74] Vicente J, Laguna P, Bartra A, Bailón R (2016) Drowsiness detection using heart rate variability. Med Biol Eng Comput 54(6):927–937. DOI: 10.1007/s11517-015-1448-7

[75] Wang Y, Wei S, Zhang S, Zhang Y, Zhao L, Liu C, Murray A (2018) Comparison of time-domain, frequency-domain and non-linear analysis for distinguishing congestive heart failure patients from normal sinus rhythm subjects. Biomed Signal Process Control 42:30–36. DOI: 10.1016/j.bspc.2018.01.001

[76] Wessel N, Malberg H, Bauernschmitt R, Kurths J (2007) Nonlinear methods of cardiovascular physics and their clinical applicability. Int J Bifurcation Chaos 17(10):3325. DOI: 10.1142/S0218127407019093

[77] Xie HB, Guo YJ, Zheng YP (2010) Fuzzy approximate entropy analysis of chaotic and natural complex systems: detecting muscle fatigue using electromyography signals. Ann Biomed Eng 38(4) 1483–1496. DOI: 10.1007/s10439-010-9933-5

[78] Yu S-N, Lee M-Y (2012) Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability. Comput Methods Programs Biomed 108(1):299–309. DOI: 10.1016/j.cmpb.2011.12.015

[79] Yu S-N, Lee M-Y (2012) Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. Comput Biol Med 42:816–825. DOI: 10.1016/j.compbiomed.2012.06.005

[80] Yu L, Liu H (2003) Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. In: Proc 20th Int Conf Mach Learn (ICML-2003), pp. 856–863.

[81] Zhang X-S, Zhu Y-S, Thakor NV, Wang Z-Z (1999) Detecting ventricular tachycardiaand fibrillation by complexity measure. IEEE Trans Biomed Eng 46(5):548–555. DOI: 10.1109/10.759055