

Discrimination power of long-term heart rate variability measures for Chronic Heart Failure detection

Paolo Melillo^{*}, Roberta Fusco^{*}, Mario Sansone^{*}, Marcello Bracale^{*}, Leandro Pecchia^{*}

^{*} *Department of Biomedical, Telecommunication and Electronic Engineering, DIBET, University Federico II of Naples, Via Claudio 21, 80100 Napoli, Italy*

Corresponding Author: Leandro Pecchia

e-mail: leandro.pecchia@unina.it

Copyright © Springer. The final publication is available at:

<http://www.springerlink.com/content/c5739h26192h7762/>

This article should be cited as follows:

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 Bio Eng Comput 49 (1):67-74. doi:10.1007/s11517-010-0728-5

Abstract

The aim of this study was to investigate the discrimination power of standard long-term Heart Rate Variability (HRV) measures for the diagnosis of Chronic Heart Failure (CHF).

We performed a retrospective analysis on 4 public Holter databases, analyzing the data of 72 normal subjects and 44 patients suffering from CHF. To assess the discrimination power of HRV measures, we adopted an exhaustive search of all possible combinations of HRV measures and we developed classifiers based on Classification and Regression Tree (CART) method, which is a non-parametric statistical technique.

We found that the best combination of features is: Total spectral power of all NN intervals up to 0.4 Hz (TOTPWR), square Root of the Mean of the Sum of the Squares of Differences between adjacent NN intervals (RMSSD) and Standard Deviation of the Averages of NN intervals in all 5-minute segments of a 24-hour recording (SDANN). The classifiers based on this combination achieved a specificity rate and a sensitivity rate of 100.00% and 89.74% respectively.

Our results are comparable with other similar studies, but the method we used is particularly valuable because it provides an easy to understand description of classification procedures, in terms of intelligible “if ... then ...” rules. Finally, the rules obtained by CART are consistent with previous clinical studies.

Keywords

Heart Rate Variability (HRV), Chronic Heart Failure (CHF), Classification and regression tree (CART)

1. Introduction

Heart Rate Variability (HRV) is the variation over time of the period between consecutive heartbeats (RR intervals)[30] and is a non-invasive measure commonly used to assess the influence of the autonomic nervous system (ANS) on the heart[23]. HRV has been widely studied in patients suffering from Chronic Heart Failure (CHF)[1-3, 5, 10-11, 13, 17, 19, 27, 32].

CHF is a patho-physiological condition in which an abnormal cardiac function is responsible for the failure of the heart to pump blood as required by the body. CHF has been shown to be degenerative and age related[15].

The majority of studies in the literature used HRV measures for the prognosis of the disease, in particular as a predictor of the risk of mortality [2, 10-11, 17, 19, 32, 36]. A small number of studies[3, 13] focused on using HRV measures for CHF diagnosis. For instance, Asyali [3] studied the discrimination power of long-term HRV measures (time-domain and FFT-based frequency domain) in order to distinguish normal subject from CHF patients. Using linear discriminant analysis and a Bayesian Classifier he obtained sensitivity and specificity rates of 81.82% and 98.08% respectively. Isler et al. [13] investigated the discrimination power of short-term HRV measures, including wavelet entropy. In this study, they achieved the best performance using Genetic Algorithms and k-Nearest Neighbour Classifier, resulting in a sensitivity rate of 100.00% and a specificity rate of 94.74%. Although this study reached interesting results, the classifier proposed by Isler [13] relied on complex features and rules which are difficult for clinicians to interpret. In both these studies[3, 13] the classifiers were developed using a subset of the dataset adopted in the current study.

As regards clinical practice, international guidelines on diagnosis and management of heart failure[15] showed evidence that conventional 12-lead electrocardiogram (ECG) should not form the primary basis for determining the specific cardiac abnormality responsible for the development of CHF, because of low sensitivity and specificity.

The method we used to investigate the power of standard long-term HRV measures in distinguishing CHF patients from normal subjects is Classification and Regression Tree (CART). CART, developed by Breiman et al. [6], is widely used in several applications of pattern recognition especially for image processing and for medical diagnosis[8]. The CART algorithm iteratively splits the data set, according to a criterion that maximizes the separation of the data, producing a tree-like decision structure[6].

CART was applied to HRV measures for other investigations, such as for the diagnosis of Obstructive Sleep Apnea Syndrome[31], and for the analysis of the relationship between HRV and the menstrual cycle in healthy young women [34]. We have adopted CART in previous studies, to investigate discrimination power of short-term HRV features[28-29] in distinguishing CHF patients from normal subjects and in assessing CHF severity. To the best of the authors' knowledge, CART has not yet been applied to long-term HRV analysis for CHF diagnosis. We chose the CART method, as it provides a model of human understandable "if ... then" rules which can be easy to read and to interpret. This is crucial in medical applications[7], in which the physician is personally responsible of the diagnosis.

In this paper we used CART in order to:

- 1) define the "optimal" classifier, based on an understandable set of rules for CHF detection;
- 2) find the "optimal" combination of standard long-term HRV measures for CHF detection.

Among the classifiers achieving a good performance, we chose as "optimal" the one which demonstrated simultaneously the lowest estimated misclassification probability and the lowest degree of overfitting.

We performed a retrospective analysis of four Holter monitor public databases, and used only open source and validated HRV toolkit software in order to allow other scientists to reproduce our results. We adopted the exhaustive search method to evaluate all the possible combinations of HRV measures and in order to develop and test a classifier for each combination of measures.

2 Methods

2.1 Data

We analyzed 116 nominal 24-hour records, from 72 normal subjects and 44 suffering from CHF. The overall dataset consisted of 54 men, 43 women and 19 unknown-gender subjects, aged 20 to 79 (55 ± 14 years). The normal subjects were 35 men and 37 women, aged 20 to 76 (55 ± 16 years). The CHF subjects were 19

men, 6 women and 19 unknown-gender subjects, aged 22 to 79 (56 ± 11 years).

The data for normal subjects were retrieved from the Normal Sinus Rhythm RR Interval Database [9] and from the MIT-BIH Normal Sinus Rhythm Database [9]. The former includes RR intervals extracted from 24-hour ECG recordings from 30 men and 24 women, aged 29 to 76 (61 ± 12 years). The latter includes long-term ECG recordings from 5 men and 13 women, aged 20 to 50 (61 ± 8 years). The data for the CHF group were retrieved from the Congestive Heart Failure RR Interval Database[9] and from the BIDMC Congestive Heart Failure Database [9]. The former includes RR intervals extracted from 24-hour ECG-Holter recordings of 8 men, 2 women, and 19 unknown-gender subjects, aged 34 to 79 (55 ± 11 years). The latter database includes long-term ECG recordings from 11 men and 4 women, aged 22 to 71 (56 ± 11 years), with severe congestive heart failure.

All the records are provided with beat annotations obtained by automated analysis with manual review and correction, with the exception of beat annotations from the BIDMC Congestive Heart Failure Database which were not manually corrected. All the original ECG records were digitized at 128 samples per second, with the exception of the records from the BIDMC Congestive Heart Failure Database, which were sampled at 250 samples per second.

2.2 Long-term HRV measures

We performed standard long-term HRV analysis on nominal 24-hour recordings according to International Guidelines[23].

The series of normal to normal (NN) beat intervals were obtained from the beat annotation files of the selected four databases and the NN/RR ratio was computed as the fraction of total RR intervals classified as normal-to-normal (NN) intervals. This ratio has been used as a measure of data reliability, excluding records with a ratio less than a threshold. Thresholds of 80% [9] and 90%[3] were proposed. We chose a threshold of 85%, as it was a satisfactory trade-off between numbers of included subjects and quality of NN signals. Using this technique, 6 records were excluded (5 CHF and 1 normal) and the final dataset consisted of 110 subjects: 71 healthy people and 39 CHF patients.

All the computed basic time- and frequency-domain HRV measures were widely used in the literature [23, 30]. A number of standard statistical time-domain HRV measures are calculated: AVerage of all NN intervals (AVNN), Standard Deviation of all NN intervals (SDNN), Standard Deviation of the Averages of NN intervals in all 5-minute segments of a 24-hour recording (SDANN), mean of the Standard Deviations of NN intervals in all 5-minute segments of a 24-hour recording (SDNN IDX), square Root of the Mean of the Sum of the Squares of Differences between adjacent NN intervals (RMSSD), percentage of differences between adjacent NN intervals that are longer than 50 ms (pNN50). Moreover, percentage of

differences between adjacent NN intervals that are longer than 12 ms (pNN12) was computed because Mietus [25] showed that among pNNx measures pNN12 can provide the maximum separation between normal subjects and CHF patients.

The frequency-domain HRV measures rely on the estimation of power spectral density (PSD). Several methods were proposed in literature in order to estimate PSD of RR intervals [12, 23, 26, 30]. In this study, we estimated PSD both by Welch's averaged modified periodogram[35] and by Lomb-Scamble periodogram [22]. For the Welch's periodogram, the NN interval was first interpolated with cubic spline interpolation at 4 Hz. The interpolated series was divided into overlapping segments of length 4000 points and each segment was Hanning windowed. The overlap was chosen to be 1200 points[3]. After PSD estimation, six standard frequency-domain HRV measures were calculated: total spectral power of all NN intervals up to 0.4 Hz (TOTPWR), between 0 and 0.003 Hz (ULF), between 0.003 and 0.04 Hz (VLF), between 0.04 and 0.15 Hz (LF), between 0.15 and 0.4 Hz (HF), ratio of low to high frequency power (LF/HF). Further in the paper, we will refer Welch-based measures and the Lomb-Scamble-based ones by using the subscript W and LS, respectively. For instance, TOTPWR_W refers to the estimation of TOTPWR computed by using Welch periodogram, while TOTPWR_{LS} refers to the one obtained by Lomb-Scamble periodogram.

2.3 Feature selection

CART resembles a stepwise feature selection, as at each splitting it tries to obtain the most relevant information from the part of the space it is working on. However, one feature may not be included in the final tree because its effect was masked by other variables. In order to deal with masking and find the best subset of features, we adopted the so-called exhaustive search method[14], investigating the predictive value of all the possible combinations of k out of N features (with k from 1 to N). Since the number of features N is 13, we investigated $2^{13}=8092$ subsets of features, training and testing the same number of classification trees, as further discussed. In order to facilitate comparisons of our results with previous studies, we performed the feature selection twice: once using frequency-domain features measured by Welch periodogram; once using frequency-domain features measured by Lomb-Scamble periodogram.

2.4 Optimal classification tree for each combination of features

The CART method consists of two steps: tree growing and tree pruning [6]. In the former step, the tree grows by selecting from all the possible splits, those which generate the "purer" child nodes, where the purest node is the one containing elements of only one class. The outcome of this step is referred to as the large-tree (LT). In the latter step, the LT is pruned according to a minimal cost-complexity function, which relies on the

number of nodes and the misclassification probability. The outcome of this step is the best sub-tree (BST) which achieves the lowest value of the cost-complexity function.

In the tree growing, among different functions that have been defined for the measure of the impurity of each node t [6], we adopted the Gini index criterion[6], which, for binary classification, can be computed as follows:

$$\text{Gini index}(t) = 1 - p^2(i|t) - p^2(j|t); \quad (1)$$

where " t " is the considered node, " i " and " j " are the two class labels, $p(i|t)$ and $p(j|t)$ are the conditional probability of observing at node " t " a subject belonging to the class " i " or " j ", respectively.

In the tree pruning, CART searches among all the sub-trees, selecting the Best Sub-Tree (BST) as the one that minimizes a cost-complexity function, $R_\alpha(T)$, which is a linear combination of misclassification probability, $R(T)$, and tree complexity, $|T|$, defined as its number of nodes (see eq.2) [6].

$$R_\alpha(T) = R(T) + \alpha|T| \quad (2)$$

Where α is a non negative parameter. We estimated $R(T)$ twice: once as mean on several testing subsamples, using a Cross-Validation (CV) technique; once on all the sample, using a Re-Substitution (RS) technique. The former, named further $R(T)_{CV}$, was used to stop the tree pruning (equation 2), the later, named further $R(T)_{RS}$, was used to estimate the degree of overfitting (see eq. 3).

We estimated $R(T)_{CV}$ by the 10-fold-cross-validation (10-CV) procedure [33], which consists of randomly dividing the dataset into 10 subsamples, with almost equal size and same class proportion. The first step of the 10-CV consists of excluding a subsample (testing) in turn and developing a tree with the remaining 9 subsamples (training). The second step of the 10-CV consists of using the testing subsample to estimate its misclassification probability as the ratio between the number of misclassified cases and the total number of cases in the testing subsample. From this two step we have 10 trees, using the same combination of features, and for each a misclassification probability $r(t)_i$. $R(T)_{CV}$ is estimated as the average over these ten $r(t)_i$. Moreover, we computed the standard error for $R(T)_{CV}$, named further $SE_{R(T)_{CV}}$. Further details about minimal cost-complexity pruning can be found in Breiman [6]. Finally, $R(T)_{RS}$ was estimated using all the sample to train and test the tree.

2.5 Optimal feature combination

For each subset of features, we repeated the steps described in section 2.4, finding the BST. We estimated the overfitting of each BST by introducing a normalized measure of the degree of overfitting OvM, which compares $R(t)_{CV}$ with $R(t)_{RS}$ [16]:

$$OvM = \frac{R(T)_{CV} - R(T)_{RS}}{SE_{R(T)_{CV}}} \quad (3)$$

If OvM is higher than 1, we assumed that the tree overfitted the training data.

We chose as the best subset of features the one, or the ones, whose BST achieved a value of OvM lower than 1 and a value $R(t)_{cv}$ not higher than 10%. Finally, for the selected classifiers we computed a metric of performance measures in order to compare our system with others, using the formulae reported in Table 1.

Table 1 Binary Classification Performance Measures

Measure	Abbreviation	Formula
Accuracy	Acc	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	Pre	$\frac{TP}{TP+FP}$
Sensitivity	Sen	$\frac{TP}{TP+FN}$
Specificity	Spe	$\frac{TN}{TN+FP}$
Area Under the Curve	AUC	$AUC = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right)$

TP: the number of CHF patients detected

TN: number of normal subject detected

FP: number of normal subject incorrectly labelled as CHF

FN: number of CHF patients incorrectly labelled as normal.

2.6 Implementation details

The HRV extraction and analysis was performed using PhysioNet's HRV Toolkit[9]. We chose this toolkit as it is an open source and a rigorously validated package. In particular, the implementation of the LS periodogram adopted in this paper is the lomb.c program available from Physionet[20].

The feature selection and classification were performed by in-house software developed in MATLAB version R2009b (The MathWorks Inc., Natick, MA). In particular CART was implemented by utilizing the methods and construction of the MATLAB class *classregtree*.

3 Results

The optimal features are those whose BST achieved $R(t)_{cv} \leq 10\%$ and $OvM < 1$. these features are reported in Table 2 and their BSTs are shown in Figure 1.

Table 2 Classification Performance Measurements of the CART classifiers based on single HRV feature with $R(t)_{cv}$ not higher than 10% and OvM lower than 1

$R(t)_{cv}$	$SD_{R(t)_{cv}}$	$R(t)_{res}$	features selected by CART
7.27%	2.41%	6.36%	SDNN
8.18%	2.52%	6.36%	SDANN
9.09%	2.62%	8.18%	TOTPW _w

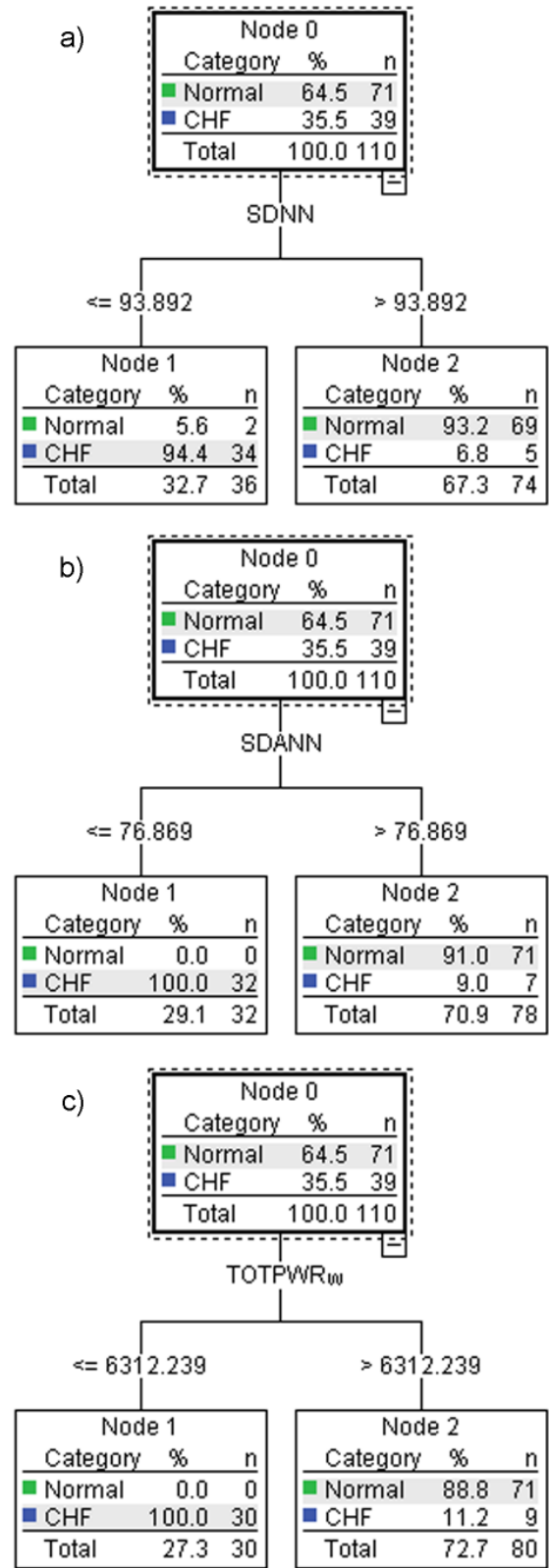


Fig. 1 The final model tree for the following HRV features: a) SDNN, b) SDANN, c) TOTPW_w

Table 3 reports the optimal combinations of features and Figure 2 shows their BSTs.

In Fig. 1 and Fig. 2 the paths from the first node to each terminal one are a graphical representation of a set of “if ... then” rules. For instance, the path to the terminal node 2 in the Fig. 2a can be read as: “if TOTPW_{LS} is higher than 8271.86 ms² and RMSDD is

higher than 15.62 ms, then the subject is classified as normal”.

Table 3 Classification Performance Measurements of the CART classifiers based on combination of HRV features with $R(t)_{cv}$ not higher than 10% and OvM lower than 1

$R(t)_{cv}$	$SD_{R(t)cv}$	$R(t)_{res}$	features selected by CART
5.46%	2.10%	3.64%	TOTPWR _{LS} , RMSSD, SDANN
10.00%	2.64%	9.09%	VLF _{LS} , pNN12

In the models shown in Fig. 1, if the selected feature (SDNN; SDANN; TOTPWR_W) was lower than a threshold (93.892 ms; 76.869 ms; 6,313.239 ms², respectively), the subject was classified as a CHF patient, otherwise as a normal subject.

In the model shown in Fig. 2a, the initial variable selected by CART (at node 1 split) was TOTPWR_{LS}. The subjects whose TOTPWR_{LS} is lower than 8271.86 ms² were correctly classified as CHF patient. CART selected RMSSD for the second node split. In this node split, the subjects whose RMSSD were higher than 15.62 ms were classified as normal. This splitting determined the four false negatives of the classifier. A final classification split is based on SDANN, that is, if it is lower than 106.71 ms, the subject is classified as CHF patient, otherwise as a normal subject.

In the model shown in Fig. 2b, the initial variable selected by CART (at node 1 split) was VLF_{LS}. The subjects whose VLF_{LS} is lower than 558.048 ms² were correctly classified as CHF patient. CART selected pNN12 for the second node split. In this node split, the subjects whose pNN12 were lower than 36.466 were classified as CHF patients, the others as normal subject. This splitting determined the nine false negatives and the one false positive of the classifier.

Table 4 shows the performance measures of the BSTs and of two classifiers proposed by other authors [3, 13].

Among those combinations of features, TOTPWR_{LS}, RMSSD and SDANN achieved the best performance in terms of the estimated probability of misclassifying any future cases. Its BST achieved specificity and sensitivity rates of 100% and 89.74%.

4 Discussion

In this study, we investigate the discrimination power of standard long-term Heart Rate Variability (HRV) measures for the diagnosis of Chronic Heart Failure. We analyzed single features and all their possible combinations.

It was indicated that the combinations of standard long-term HRV measures TOTPWR_{LS}, RMSSD, and SDANN enable distinguishing normal subjects from CHF patients with specificity and sensitivity rates of 100% and 89.74% respectively.

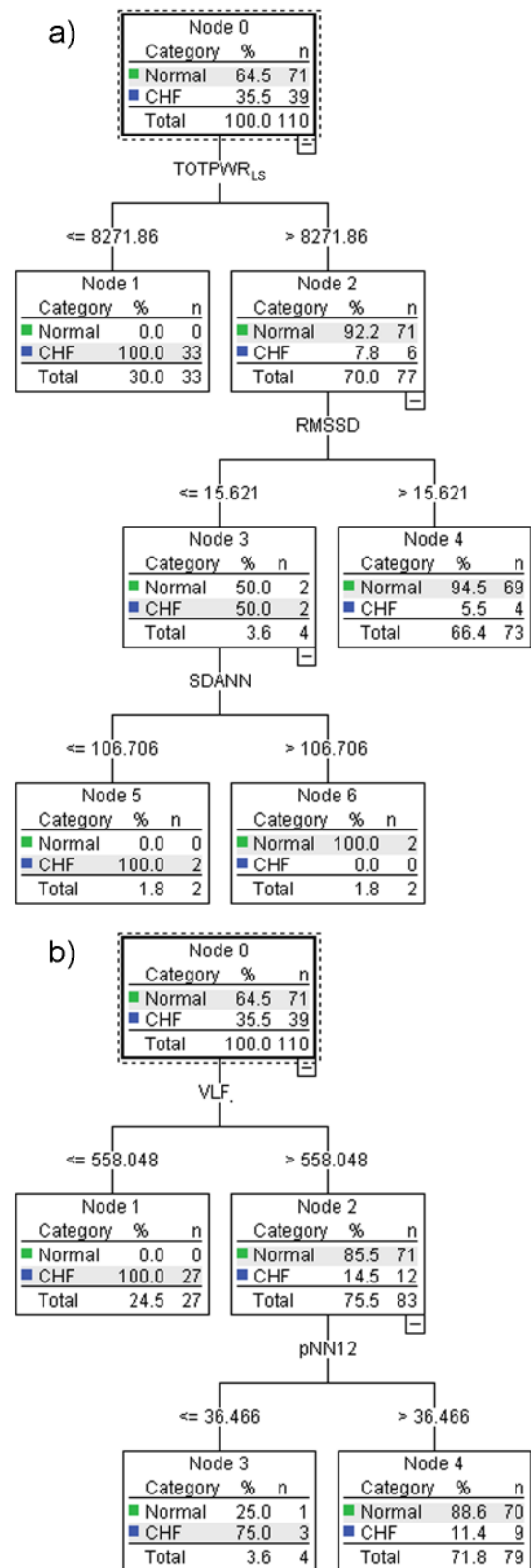


Fig. 2 The final model tree for the following combinations of HRV features:

- a) TOTPWR_{LS}, RMSSD, SDANN;
- b) VLF_{LS} and pNN12

The combination of pNN12 and VLF confirms the discrimination power of pNN12 proved by Mietus [25], but the performance of its BST is lower as shown in Table 3.

Table 4 Classification Performance Measurements of the classifiers proposed in the current study and those proposed in previously published papers.

Classifier based on single feature	TP #	FN #	TN #	FP #	ACC %	PRE %	SEN %	SPE %	AUC %
SDNN	34	5	69	2	93.64	94.44	87.18	97.18	92.18
SDANN	32	7	71	0	93.64	100.0	82.05	100.0	91.03
TOTPW _w	26	9	71	0	91.51	100.0	74.29	100.0	87.14
Asyali [3] (based on SDNN)	18	4	51	1	93.24	94.74	81.82	98.08	89.95
Classifier based on combination of features	TP #	FN #	TN #	FP #	ACC %	PRE %	SEN %	SPE %	AUC %
TOTPW _{LS} , RMSSD, SDANN	35	4	71	0	96.36	100.0	89.74	100.0	94.87
VLF _{LS} , pNN12	26	9	70	1	90.57	96.30	74.29	98.59	86.44
Isler [13] *	29	0	51	3	96.39	90.63	100.0	94.44	97.22

*based on short-term HRV measures

The sets of rules of the proposed models are clinically consistent, even if CART did not use any medical a priori knowledge. In fact, the main clinical result of this research is that terminal node classifying as CHF patients are on the left, therefore revealing lower values of the splitting features for CHF patients. This is coherent with the results showed by Bigger [5], Musialik-Lydka [27] and Arbolishvili [1], who stated that standard long-term HRV measures were significantly depressed in CHF patients, compared with normal subjects. It should be emphasized that the findings of Bigger [5] and Arbolishvili [1] were obtained adopting different methods for power spectral density estimation, while Musialik-Lydka [27] considered only time-domain HRV measures.

With regard to the performance of the classifiers in this study, they can be compared with a few previously published studies [3, 13], which developed a classifier in order to distinguish CHF patients from normal subject, as reported in Table 4. In both these studies the classifiers were developed using a smaller dataset than in the current study, as they used only the MIT-BIH Normal Sinus Rhythm Database [9] and the Congestive Heart Failure RR Interval Database [9].

The performance measures of our classifiers are higher than those of Asyali's classifier [3], which used HRV long term measures. Asyali [3] identified SDNN and SDANN as the HRV measures with the highest class separation power and the results in the current study confirm this identification. Moreover, our study showed that TOTPW_w is the third measure for separation power (see Table 2). This is not in line with Asyali findings [3], which showed that the TOTPW_w discrimination power is the second-last. This could be because Asyali used Fisher's Linear Discriminant Analysis (LDA) without proving normal distribution of HRV features. In fact, LDA is strongly affected by non-normality of data [18] and for that reason is expected to provide less accurate information about non-normal measures, such as TOTPW [4].

The performance measures of our classifier are lower than those of Isler's classifier [13], which used HRV short-term measures. The best combination of features selected by Isler consists of 8 features including sex, FFT-based measures, LS periodogram measures and wavelet entropy measures with a sensitivity rate of 100.00% and a specificity rate of 94.74% [13]. The higher sensitivity rate may be because of the discrimination power of wavelet entropy measures, which have not been considered in this study because they are not standard HRV measures, presumably too complex to understand for most clinicians. Moreover, Isler's classifiers [13] were based on at least 8 features.

As regards the chosen method, this improved intelligibility in comparison to Isler's one [13] and minimized the risk of overfitting in comparison to Asyali's one [3]. In fact, Isler [13] adopted k-nearest-neighbor (KNN) classifiers, which lack the property of the interpretability of induced knowledge [21]. While, Asyali [3] developed a Bayesian Classifier and the classification is based on the following rule: "a subject with an SDNN value higher (lower) than tan threshold t_{opt} is classified as normal (abnormal)". This rule is the same of the BST in fig. 2a, with a slightly different threshold: 93.89 for CART classifier, 91.82 for Bayes classifier. However, Asyali computed the threshold on the whole dataset and provided no information about probability of misclassification, as he did not use cross-validation approach nor independent test set.

Our study had the following limitations: a small and unbalanced dataset, the differences in the sampling frequency of ECG recordings and the different extraction procedures of NN intervals. As concerns the sampling frequency of ECG, it should be remembered that the finite sampling frequency introduce an error in the RR interval measurement as shown by Merri [24]. However, a sampling rate of 128 Hz, which is the lowest sampling rate of the records used in this paper, has been found to be accurate enough to locate the R-peaks and hence compute HRV [23]. In future research we are planning to develop a similar study on a larger

dataset of ECG recordings digitalized at the same sampling frequency and annotated with a stated procedure.

We concluded that the long-term HRV measures enable CHF patients to be distinguished from normal subjects confirming results of previous studies. In particular, the combination TOTPWR_{LS}, RMSSD, SDANN allows detecting CHF patients with a precision and a specificity of 100.00%, an accuracy of 96.36%, a sensitivity of 89.74% and an estimated misclassification probability of 5.46%.

Our results are consistent with the consensus that depressed HRV values are good indicators of CHF. Furthermore, we estimated numerical threshold values for a set of a few standard HRV measures, arranged in a decision tree, which is fully human intelligible.

References

1. Arbolishvili GN, Mareev VY, Orlova YA, Belenkov YN (2006) Heart rate variability in chronic heart failure and its role in prognosis of the disease. *Kardiologiya* 46 19:4-11
2. Aronson D, Mittleman MA, Burger AJ (2004) Measures of heart period variability as predictors of mortality in hospitalized patients with decompensated congestive heart failure. *Am J Cardiol* 93 54:59-63
3. Asyali MH (2003) Discrimination power of long-term heart rate variability measures. In: *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, 200-203 Vol.201
4. Bigger JT, Fleiss JL, Steinman RC, Rolnitzky LM, Kleiger RE, Rottman JN (1992) Correlations among time and frequency-domain measures of heart period variability 2 weeks after acute myocardial-infarction. *Am J Cardiol* 69 5693:891-898
5. Bigger JT, Fleiss JL, Steinman RC, Rolnitzky LM, Schneider WJ, Stein PK (1995) Rr variability in healthy, middle-aged persons compared with patients with chronic coronary heart-disease or recent acute myocardial-infarction. *Circulation* 91 1095:1936-1943
6. Breiman L (1984) Classification and regression trees. Wadsworth International Group, Belmont, Calif.
7. Cios KJ, Moore GW (2002) Uniqueness of medical data mining. *Artif Intell Med* 26 1282:1-24
8. Esposito F, Malerba D, Semeraro G (1997) A comparative analysis of methods for pruning decision trees. *Ieee T Pattern Anal* 19 1304:476-491
9. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, 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* 101 557:e215-220
10. Guzzetti S, Magatelli R, Borroni E, Mezzetti S (2001) Heart rate variability in chronic heart failure. *Autonomic Neuroscience-Basic & Clinical* 90 64:102-105
11. Hadase M, Azuma A, Zen K, Asada S, Kawasaki T, Kamitani T, Kawasaki S, Sugihara H, Matsubara H (2004) Very low frequency power of heart rate variability is a powerful predictor of clinical prognosis in patients with congestive heart failure. *Circulation Journal* 68 49:343-347
12. Ihlen EAF (2009) A comparison of two hilbert spectral analyses of heart rate variability. *Medical & Biological Engineering & Computing* 47 6897:1035-1044
13. Isler Y, Kuntalp M (2007) Combining classical hrv indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine* 37 15:1502-1510
14. Jain AK, Duin RPW, Jianchang M (2000) Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22 183:4-37
15. Jessup M, Abraham WT, Casey DE, Feldman AM, Francis GS, Ganiats TG, Konstam MA, Mancini DM, Rahko PS, Silver MA, Stevenson LW, Yancy CW, Hunt SA, Chin MH, Jessup M, Michl K, Oates JA, Smith SC, Jacobs AK, Buller CE, Creager MA, Ettinger SM, Krumholz HM, Kushner FG, Lytle BW, Nishimura RA, Page RL, Tarkington LG, Yancy CW, Lewin JC, May C, Bradfield L, Stewart MD, Keller S, McDougall A, Brown N, Whitman GR (2009) 2009 focused update: Accf/aha guidelines for the diagnosis and management of heart failure in adults a report of the american college of cardiology foundation/american heart association task force on practice guidelines developed in collaboration with the international society for heart and lung transplantation. *Journal of the American College of Cardiology* 53 145:1343-1382
16. Khoshgoftaar TM, Allen EB (2001) Controlling overfitting in classification-tree models of software quality. *Empirical Software Engineering* 6 5698:59-79
17. Kruger C, Lahm T, Zugck C, Kell R, Schellberg D, Schweizer MWF, Kubler W, Haass A (2002) Heart rate variability enhances the prognostic value of established parameters in patients with congestive heart failure. *Zeitschrift Fur Kardiologie* 91 58:1003-1012
18. Krzanowski WJ (1977) The performance of fisher's linear discriminant function under non-optimal conditions. *Technometrics* 19 5744:191-200
19. La Rovere MT, Pinna GD, Maestri R, Mortara A, Capomolla S, Febo O, Ferrari R, Franchini M, Gnemmi M, Opasich C, Riccardi PG, Traversi E, Cobelli F (2003) Short-term heart rate variability strongly predicts sudden cardiac death in chronic heart failure patients. *Circulation* 107 57:565-570
20. Laguna P, Moody GB, Mark RG (1998) Power spectral density of unevenly sampled data by least-square analysis: Performance and application to heart rate signals. *Ieee Transactions on Biomedical Engineering* 45 559:698-715
21. Lavrač N (1999) Machine learning for data mining in medicine In: *Artificial intelligence in medicine*, Springer Berlin / Heidelberg, 47-62
22. Lomb (1976) Least-squares frequency analysis of unequally spaced data (in astronomy) *Astrophysics and Space Science* 39 560:447-462

23. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, Schwartz PJ (1996) Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Eur Heart J* 17 134:354-381
24. Merri M, Farden DC, Mottley JG, Titlebaum EL (1990) Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *Biomedical Engineering, IEEE Transactions on* 37 3235:99-106
25. Mietus JE, Peng CK, Henry I, Goldsmith RL, Goldberger AL (2002) The pnnx files: Re-examining a widely used heart rate variability measure. *Heart* 88 59:378-380
26. Mitov IP, Daskalov IK (1998) Comparison of heart rate variability spectra using generic relationships of their input signals. *Med Biol Eng Comput* 36 3170:573-580
27. Musialik-Lydka A, Sredniawa B, Pasyk S (2003) Heart rate variability in heart failure. *Kardiol Pol* 58 1114:10-16
28. Pecchia L, Melillo P, Bracale M (2010) Remote health monitoring of heart failure with data mining via cart method on hrv features. *Biomedical Engineering, IEEE Transactions on* PP 6981:1-1
29. Pecchia L, Melillo P, Sansone M, Bracale M (2010) Discrimination power of short-term heart rate variability measures for chf assessment. *Information Technology in Biomedicine, IEEE Transactions on* PP 6983:1-1
30. Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS (2006) Heart rate variability: A review. *Med Biol Eng Comput* 44 3151:1031-1051
31. Roche F, Gaspoz J-M, Court-Fortune I, Minini P, Pichot V, Duverney D, Costes F, Lacour J-R, Barthelemy J-C (1999) Screening of obstructive sleep apnea syndrome by heart rate variability analysis. *Circulation* 100 121:1411-1415
32. Smilde TDJ, van Veldhuisen DJ, van den Berg MP (2009) Prognostic value of heart rate variability and ventricular arrhythmias during 13-year follow-up in patients with mild to moderate heart failure. *Clinical Research in Cardiology* 98 4:233-239
33. Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)* 36 593:111-147
34. Vallejo M, Marquez MF, Borja-Aburto VH, Cardenas M, Hermosillo AG (2005) Age, body mass index, and menstrual cycle influence young women's heart rate variability - a multivariable analysis. *Clinical Autonomic Research* 15 126:292-298
35. Welch P (1967) The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *Audio and Electroacoustics, IEEE Transactions on* 15 5745:70-73
36. Wessel N, Voss A, Kurths J, Schirdewan A, Hnatkova K, Malik M (2000) Evaluation of renormalised entropy for risk stratification using heart rate variability data. *Med Biol Eng Comput* 38 3165:680-685