IOP Publishing Journal Title
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

Detection of subjects with Ischemic Heart Disease by using machine learning technique based on Heart Rate Total Variability parameters

Accardo Agostino, Silveri Giulia, Merlo Marco, Restivo Luca, Ajčević Milos, Sinagra Gianfranco
[1] Department of Engineering and Architecture, University of Trieste, Trieste, Italy
[2] Cardiovascular Department, University of Trieste, Trieste, Italy

E-mail: giulia.silveri@phd.units.it

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

Objective: Ischemic heart disease (IHD), in its chronic stable form, is a subtle pathology due to its silent behavior before developing in unstable angina, myocardial infarction or sudden cardiac death. The clinical assessment is based on typical symptoms and finally confirmed, invasively, by coronary angiography. Recently, heart rate variability (HRV) analysis as well as some machine learning algorithms like Artificial Neural Networks (ANNs) were used to identify cardiovascular arrhythmias and, only in few cases, to classify IHD segments in a limited number of subjects. The goal of this study was the identification of the ANN structure and the HRV parameters producing the best performance to identify IHD patients in a non-invasive way, validating the results on a large sample of subjects. Moreover, we examined the influence of a clinical non-invasive parameter, the left ventricular ejection fraction (LVEF), on the classification performance. **Approach:** To this aim, we extracted several linear and non-linear parameters from 24h RR signal, considering both normal and ectopic beats (Heart Rate Total Variability), of 251 normal and 245 IHD subjects, matched by age and gender. ANNs using several different combinations of these parameters together with age and gender were tested. For each ANN, we varied the number of hidden neurons from 2 to 7 and simulated 100 times changing randomly training and test dataset. **Main Results:** The HRTV parameters showed significant greater variability in IHD than in normal subjects. The ANN applied to meanRR, LF, LF/HF, Beta exponent, SD2 together with age and gender reached a maximum accuracy of 71.8% and, by adding as input LVEF, an accuracy of 79.8%. **Significance:** The study provides a deep insight into how a combination of some HRTV parameters and LVEF could be exploited to reliably detect the presence of subjects affected by IHD.

Keywords: Ischemic heart disease, Heart Rate Variability, Artificial Neural Networks, Non-linear analysis, Left ventricular ejection fraction

1. Introduction

Ischemic heart disease (IHD) is a pathological condition characterized by an imbalance between myocardial oxygen supply and demand, mainly due to coronary artery atherosclerosis [1]. Although the narrowing can be caused by a blood clot or by constriction of the blood vessel, most often it is caused by buildup of plaque. Generally, this disease is silent, showing no symptoms but can become unstable due to plaque rupture or erosion causing angina, myocardial infarction or sudden cardiac death [2].

Current clinical guidelines suggest using several clinical diagnostic techniques to confirm the disease, such as those based on single photon emission computed tomography, radionuclide myocardial perfusion imaging, positron emission tomography, cardiac magnetic resonance imaging and multi slice computed tomography. All these techniques present invasiveness and at least one limitation, such as low sensitivity, specificity, high cost and/or long examination time [3].

On the other hand, the basic testing, in patients with suspected IHD, includes laboratory biochemical test, ambulatory electrocardiogram (ECG) monitoring, stress test ECG, resting ECG and resting echocardiography. Ambulatory electrocardiogram is a tool that records the electrical activity of the heart over the time and it may detect silent myocardial ischemia by examining deviations in the ST segment or/and changes in T wave that are due to the disease. These ECG abnormalities could occur during normal daily activities as well as during night. However, the ECG alone could not be sufficient for the diagnosis in case of slight alteration of the repolarization, due to IHD. Furthermore, some authors [4-6] underlined that the time variation of the interval between consecutive normal heartbeats (Heart Rate Variability, HRV) due to the interaction between sympathetic and parasympathetic activities in autonomic functioning [7], presented significant differences between IHD and healthy subjects. In particular, using linear and non-linear features, they highlighted in IHD patients a decrease of the HRV parameters values in the time domain [4, 5] as well as of the normalized low frequency power, of the ratio between low and high frequency powers [6]. Moreover, Voss et al. [5] suggested that the SD1 parameter, extracted from the Poincaré plot, is a prognostic index to differentiate IHD from healthy subjects.

Nowadays, different mathematical approaches for decision support system (DSS) based on HRV [8-21] as well as on some clinical features [22,23] have been developed to identify beats and arrhythmias and to classify cardiovascular diseases like atrial fibrillation, left bundle branch block, cardiomyopathy and ventricular fibrillation. These DSS are based on widely applied linear programming classification methods like the Kth nearest-neighbours (KNN) [8], clustering [9], linear discrimination analysis (LDA) [10,11], fuzzy analysis [12-14], decision tree [15,16], support vector machine (SVM) [17,18] and artificial neural network (ANN) [19-21]. Some of these algorithms have also been used to classify subjects suffering from IHD. In particular, Kannathal et al. [14] used an adaptive neuro-fuzzy network applied to three non-linear parameters extracted from HRV (Largest Lyapunov Exponent, Spectral Entropy, Poincaré plot) obtaining an accuracy of 96%. However, the data were taken from half-hour segments of the 47 subjects belonging to MIT-BIH arrhythmia database of which only a few subjects presenting ischemic/dilated cardiomyopathy. Acharya et al. [20], using the same database and the same non-linear HRV parameters, developed a cardiac abnormalities classification system by using an ANN, obtaining an accuracy of 83.3% for identification of patients with ischemic/dilated cardiomyopathy from healthy subjects. Finally, Dua et al. [21] extracted non-linear features from HRV (by using recurrence plots, Poincaré plot, detrended fluctuation analysis, Shannon’s entropy, approximation entropy and sample entropy) and, after the application of main component analysis, used the first six most significant principal components and different classification techniques to classify 143 samples recorded on 10 IHD and 10 healthy subjects. Using ANN with 4 hidden layers, they obtained a classification accuracy of 89.5% on this small sample of subjects.

Conversely, some authors applied machine learning techniques to identify IHD patients without considering HRV parameters among their inputs. In particular, Rajeswari et al. [22], using an ANN, selected 12 clinical features able to identify 712 IHD patients. They selected age, gender, menopause, body mass index, waist circumference, systolic and diastolic blood pressure, diabetes, cholesterol, hereditary, personal habits and stress features from 17 initial parameters, obtaining an accuracy of 82.2%. In addition, Kukar et al. [23], by using some machine learning techniques such as Bayesian classifier, artificial neural network, decision tree and k-nearest neighbor method, evaluated the performance of some diagnostic methods like clinical examinations, exercise ECG testing and myocardial scintigraphy for IHD identification in order to reduce the number of subjects to be submitted to coronary angiography. They used a dataset of 327 IHD patients, obtaining with a multilayer feedforward neural network, an accuracy of 92%, reducing the number of subjects unnecessarily submitted to much more invasive and dangerous examination by 12.2%.

Finally, as a basic non-invasive clinical tool to support the diagnose of IHD, the echocardiography provides the left ventricular ejection fraction (LVEF) parameter that has been accepted as a prognostic indicator of patients with IHD [1,3]. Low values of this parameter may increase the suspicion of ischemic myocardial damage.

Although several linear and non-linear HRV parameters and noninvasive techniques individually proved to achieve a good performance in the classification of IHD beats or tracts, until now the studies using ANN or other machine learning methods have been carried out only on a very limited number of IHD patients and separately using either linear or non-linear HRV parameters.

On the other hand, although the correct procedure to examine HRV signal involves a pre-processing of the ECG by detecting the ectopic beats, providing a clean normal-to-normal inter-beat series containing information about the control systems that govern sinus node stimulation, the identification of ectopic beats represents a difficult task since these beats can have a waveform similar to the normal ones. Moreover, since in the IHD patients the ectopic beats occur more frequently than in the normal heart probably providing independent information on ischemia, in this work we evaluated the heart rate total variability (HRTV), considering RR segments containing both normal and ectopic beats. However, it is worth considering that the HRTV signal contains two main components of different origin leading to opposite information about the ischemic/healthy heart. The first component is generated by the control of the ANS on the sinus node and generally decreases in conditions of ischemic heart disease compared to normal. The second component is due to ectopic beats that most affects the IHD heart and increases in the case of ischemic heart disease compared to normal. Therefore, it is not easy to link an increase of one of these variability indexes to a healthy or ischemic heart. However, these two components are of different types and affect HRTV parameters differently. In fact, typically ectopic beats lead to a substantial increase in the squared differences of successive RR intervals (RMSSD), but a generic flat increase in spectral power density. Instead, a deep respiratory sinus arrhythmia (characteristic of a healthy ANS and heart, producing series of RR intervals with quite large beat-to-beat interval variations) leads to a generic increase in RMSSD but to a specific increase of the high frequency power. Therefore, it may be interesting to evaluate whether an ANN using as input a set of parameters extracted directly from the RR series, without pre-processing for the exclusion of ectopic beats, is able to identify patients with IHD.

Thus, the aim of this study is to assess the accuracy of several multi-layer feed forward neural networks based on linear or/and non-linear parameters extracted from HRTV together with age and gender, in order to evaluate which combination of them produces the best performance for the identification of IHD patients, validating the result on a large sample of subjects. In addition, we would examine whether the addition of a non-invasive clinical parameter, i.e. the left ventricular ejection fraction, could improve the classification performance. We select HRTV features also because they provided useful information for an early diagnosis of the IHD disease and consequently to reduce the effects of this pathology and the mortality rate.

2. Methods

2.1 Study population

The cohort of this study was composed of 681 normal subjects and 284 IHD patients, consecutively enrolled from December 2016 to October 2018 at the Cardiovascular Department, ASUGI, Trieste. In order to match subjects by gender and age, we considered a subgroup of 496 subjects consisting of 251 normal subjects (188 males aged 73±8 and 63 females aged 77±7) and 245 suffered from IHD (186 males 74±7 and 59 females aged 78±7). Normal subjects did not present neither peripheral artery disease, thyroid disorders nor history of myocardial revascularization. The diagnosis of IHD was established based on guidelines examining typical symptoms such as angina, laboratory test (increase of Troponin I), ECG repolarization abnormalities (T-wave and ST), wall motion abnormalities (hypokinesia or akinesia) detected with echocardiography or positive stress testing, and finally confirmed by coronary angiography result. Following protocol, we performed a stress test ECG only in patients with intermediate CAD risk. We considered IHD patients with chronic coronary syndromes excluding patients with the left ventricular ejection fraction lower than 50%, in such way the patients did not present systolic heart failure. The retrospective study was performed according to the Declaration of Helsinki and all patients gave written consent.

2.2 *HRTV Analysis*

All subjects underwent a 24h Holter ECG monitoring, low-pass filtering and sampling the ECG signal at 75 Hz and 200 Hz, respectively. RR intervals were automatically identified by using SyneScope Holter analysis software (Sorin Group, Italy) and analyzed by a proprietary MATLAB program. Since the original recordings started at different times, the subjects RR time series were at first aligned having as a common start time 10:00am. A pre-processing RR time series was carried out interpolating the RR intervals by cubic spline, starting from those classified by SyneScope software as normal or ectopic, substituting beats classified as artifact. The resulting signal was successively resampled at 2Hz, in order to obtain a uniformed time sampled sequence, and segments of 300s along the entire 24h were individually examined and then mediated, considering at least seven hours during daytime and other seven hours during nighttime in order to take into account the influence of the day-night alternance. In order to minimize the effects of artifacts substitution, yielding a more reliable estimation of the HRTV parameters, each segment was accepted if the percentage of artifacts was less than 5% and the duration of the longest artifact tract was shorter than 10s [24].

Since the RR signal can be mainly examined by considering the system generating it as stochastic or chaotic deterministic, several linear and non-linear parameters were used for HRTV characterization. Three types of HRTV analyses were performed, extracting features from each RR segment in time, spectral and non-linear domains. The mean value of each parameter along the 24h was considered for each subject.

In the time domain linear analysis [25], the mean of the RR intervals (meanRR), the standard deviation of the RR intervals (SDRR), the root mean square of the squared differences of successive RR intervals (RMSSD), the number of differences of successive RR intervals greater than 50ms (RR50) and the proportion of RR50 divided by the total number of RR intervals (pRR50) were calculated on each interval. In the frequency domain linear analysis [25], the power spectrum density (PSD) was computed using the periodogram method (with the Hamming window) and both the absolute (LF, HF) and normalized (LFn, HFn) powers in Low (0.041-0.15Hz) and High (0.151-0.4 Hz) Frequency spectral bands were estimated. In addition, we calculated the ratio between LF and HF (LF/HF). In the non-linear domain analysis, we considered the Poincarè plot parameters [26] that portray the nature of RR interval fluctuations, the Fractal Dimension, calculated by the Higuchi’s algorithm [27] and the beta exponent [28] both quantifying HRTV self-affinity and cardiac system complexity. The Poincaré plot was analyzed quantitatively by calculating the standard deviations referred as SD1 and SD2 and their ratio, SD1/SD2. SD1 relates to the fast beat to beat variability, SD2 describes the long-term variability of RR and SD1/SD2 represents their balance. Fractal dimension, quantifying the fractal scaling properties of short interval RR time series and the self-affinity characteristics of HRTV signal, was calculated directly in time domain while beta exponent was evaluated in frequency domain as the slope of the regression line of the log(power) versus log(frequency).

2.3 *Artificial Neural Network*

ANN represents a structure and an algorithm inspired by the highly interactive processing of the human brain and, like it, the ANN can recognize patterns, manage data and learn [29]. The decision-making process of ANN is holistic, based on the features of input patterns. In this paper, a multilayer feedforward neural network was chosen as a nonlinear classifier using the generalized back-propagation algorithm. This is a supervised learning algorithm, in which a mean square error function is defined and the learning process aims to reduce the overall system error to a minimum. This algorithm showed excellent performance on pattern recognition and on the modeling of non-linear relationships involving several variables [30].

In order to assess the performance of ANNs based only on parameters derived by non-invasive HRTV analysis for IHD classification, at first we developed and compared six different ANN structures considering the HRTV normalized parameters together with age and gender. With the aim to evaluate the power of the linear or non-linear HRTV parameters used separately and to assess if their combinations could improve performance, we separately considered as input only the linear (in time and spectral domains) parameters (ANN1) or only the non-linear parameters (ANN3) and all the fifteen parameters all together (ANN5). Successively, to reduce the number of input parameters, we evaluated the Pearson correlation coefficient *r* among all parameters and examined other two ANNs excluding either linear (ANN2) or non-linear (ANN4) parameters significantly ($p<0.05$) correlated for more than 90%. For each ANN, we varied the number of hidden neurons from 2 to 7 and selected that one producing the highest accuracy. In addition, we considered all the ANNs obtained selecting as input one or more linear parameters considered for ANN2 combined together with one or more non-linear parameters considered for ANN4. For each of the 1785 resulting ANNs, the number of hidden neurons could vary from 2 to 7. The input and hidden neurons combination producing the highest accuracy was then selected (ANN6). All the ANNs had two output nodes corresponding to the two possible classes: normal or IHD. The features, the number of inputs and the number of hidden neurons selected for each ANN are reported in Table 1.

Table 1. ANN characteristics.

	# input	# hidden neurons	input parameters
ANN1	12	3	meanRR, SDRR, RMSSD, RR50, pRR50, LF, HF, LF/HF, LFn, HFn, age, gender
ANN2	10	2	meanRR, RMSSD, RR50, LF, HF, LF/HF, LFn, age, gender
ANN3	7	3	Beta exponent, SD1, SD2, SD1/SD2, FD, age, gender
ANN4	5	3	Beta exponent, SD2, FD, age, gender
ANN5	17	3	meanRR, SDRR, RMSSD, RR50, pRR50, LF, HF, LF/HF, LFn, HFn, Beta exponent, SD1,SD2, SD1/SD2, FD, age, gender
ANN6	12	3	meanRR, RR50, pRR50, LF, HF, LF/HF, LFn, Beta exponent, SD2, age, gender
ANN12	8	4	meanRR, LF, LF/HF, Beta exponent, SD2, age, gender, LVEF

Since the usefulness of a specific clinical parameter, nominally the left ventricular ejection fraction, in the diagnosis of IHD has been demonstrated [1,3], we added this feature to the inputs of each of the first five ANNs, considering ANN7 to ANN11 new networks. Finally, as the last case (ANN12), we repeated the procedure used for the ANN6, by examining 1785 new ANNs varying, for each of them, the number of hidden neurons from 2 to 7, selecting the ANN producing the highest accuracy (Table 1).

In all cases, the training and test sizes were respectively 75% and 25% of the total number of data. The training of the neural network ended if the sum of the square errors for all data was less than 0.05 or the maximum number of training epochs (1000 iterations) was reached. Each of the twelve ANNs was simulated 100 times changing randomly the training and test dataset. The performance of each network classifier was measured through accuracy (ACC), sensitivity (SEN), specificity (SPE) and precision (PRE) characteristics (Table 2).

Table 2. Performance measurements; TP= True Positive; TN=True Negative; FN=False Negative; FP=False Positive.

Symbol	Measure	Equation
ACC	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
SEN	Sensitivity	$\frac{TP}{TP+FN}$
SPE	Specificity	$\frac{TN}{TN+FP}$
PRE	Precision	$\frac{TP}{TP+FP}$

In addition, the receiver operating characteristics curve (ROC) was used to depict trade-offs between hit rate (sensitivity) and false alarm rate (1.0-specificity) and the area under the curve (AUC) was evaluated [31].

Table 3. Comparison of HRTV between normal and IHD subjects; 25^{th}: 25^{th} percentile; 50^{th}: 50^{th} percentile; 75^{th}: 75^{th} percentile. n.s.: not significant.

Parameter	NORMAL			IHD			
	25^{th}	50^{th}	75^{th}	25^{th}	50^{th}	75^{th}	p-value
meanRR (ms)	805	894	986	855	940	1053	0.001
SDRR (ms)	38	54	87	41	66	118	0.013
RMSSD (ms)	24	48	108	36	79	165	<0.001
RR50	6	20	73	7	34	98	0.015
pRR50	0.01	0.06	0.21	0.02	0.11	0.34	0.007
LF (ms^2)	161	351	816	146	368	1038	n.s.
HF (ms^2)	93	306	1769	162	938	3608	0.001
LF/HF	0.43	0.94	2.27	0.30	0.54	1.02	<0.001
LFn	0.30	0.48	0.69	0.23	0.35	0.50	<0.001
HFn	0.31	0.52	0.70	0.50	0.65	0.77	<0.001
Beta exp (ms^2/Hz)	0.41	0.93	1.35	0.21	0.55	1.01	<0.001
SD1 (ms)	15.3	23.3	44.7	17.9	34.1	67.4	0.004
SD2 (ms)	47.5	64.2	99.5	49.1	74.1	121	n.s.
SD1/SD2	0.29	0.37	0.48	0.34	0.43	0.55	0.0001
FD	1.47	1.57	1.70	1.54	1.64	1.73	<0.001

3. Results

Table 3 shows the descriptive statistics of HRTV measures in normal and IHD subjects reported, for each group, in terms of median, 25th and 75th percentiles; the significance of difference between the two subject groups was assessed by the Wilcoxon rank sum test.

Table 4. Best classification performance of ANNs; H = Hidden neurons, SEN = Sensitivity, SPE = Specificity; PRE = Precision; ACC = Accuracy, AUC= Area Under the ROC Curve.

ANN Classifier	H	SEN %	SPE %	PRE %	ACC %	AUC %
ANN1	3	71.2	67.2	71.2	69.4	70.7
ANN2	2	80.3	62.1	70.7	71.8	71.5
ANN3	3	75.4	65.7	65.2	70.2	70.5
ANN4	3	74.2	58.6	67.1	66.9	73.7
ANN5	3	75.0	60.9	64.3	67.7	70.0
ANN6	3	83.1	56.9	63.6	69.4	74.5
ANN7	3	79.2	80.3	75.0	79.8	84.5
ANN8	2	80.0	78.1	77.4	79.0	84.7
ANN9	3	74.2	79.0	78.0	76.6	83.7
ANN10	3	69.6	83.8	78.0	77.4	84.7
ANN11	3	71.4	82.0	80.4	76.6	87.1
ANN12	4	75.4	84.1	82.1	79.8	86.0

Among time domain measures, IHD patients presented significant higher values than normal subjects while all the frequency domain parameter values were lower in IHD except HF, HFn and LF. Considering the non-linear parameters, exponent beta was statistically lower in IHD patients than in normal subjects while the Poincaré measurements and FD were significantly higher in IHD patients except for SD2 that did not present significant difference.

Table 4 presents the classification performances and the area under the curve in the test phase of all the twelve considered ANNs. Among the 100 possible combinations of data used as input in the test for each ANN, we selected that producing the highest accuracy.

Almost all the best ANNs used a number of hidden neurons of three, only ANN8 and ANN12 used either 2 or 4 neurons, respectively. The performances obtained using LVEF were always better than those without it, achieving an accuracy of 79.8% with ANN12.

Figure 1 shows the distribution values of the sensitivity, specificity, precision and accuracy obtained using the 100 different combinations of data input for the test set in the ANNs presenting the best performance in the two situations (with or without adding LVEF as input parameter). Similar quite symmetrical distributions were obtained for all the other considered ANNs in the test phase. The mean (±SD) values of the sensitivity, specificity, precision and accuracy distributions for ANN2 and ANN12 were 67.8±8, 55.7±7, 59.2±5, 60.5±4 and 65.4±7, 75.1±7, 72.0±6, 70.2±4, respectively. Slightly better results were obtained in the training phase for all the ANNs.

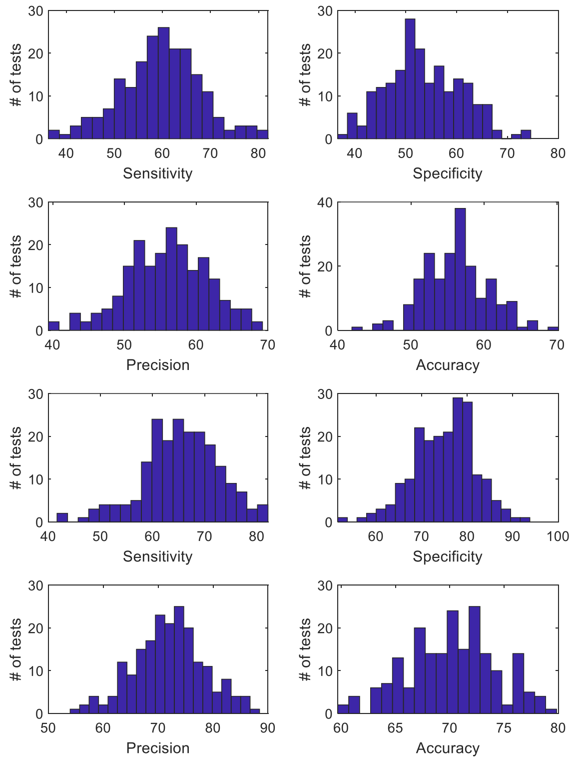

Figure 1. Distributions of the sensitivity, specificity, precision, accuracy obtained in the 100 repetitions of the test set for ANN2 (top panels) and for ANN12 (bottom panels), networks without and with LVEF, respectively.

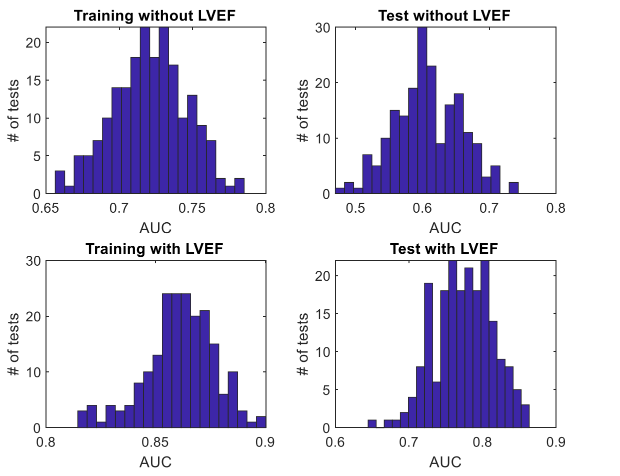

Figure 2. Distributions of the AUC(%) values obtained for the 100 different combination of the data input in the ANN2 (left panels, without LVEF) and in ANN12 (right panels, with LVEF).

Figure 2 shows the distributions of AUC obtained for ANN2 and ANN12 in the 100 repetitions of training and test phases. The mean (±1SD) values for ANN2, without LVEF, were of 71.7±2 for training and 64.6±4 for test phase, respectively and for ANN12, with LVEF, of 86.1±2 and 77.5±4 in the two phases. Finally, Figure 3 reports the ROC curves for the ANN2 and ANN12, presenting the best accuracy, during training and test phases.

4. Discussion

Patients suffered from IHD are generally asymptomatic or exhibit no typical signs and symptoms until this disease manifests itself as angina, myocardial infarction or sudden cardiac death. To reduce the effects of this disease and consequently the mortality rate, an early diagnosis of this pathology is necessary. To automate this detection task, we propose a decision support system based on ANN applied to features coming from HRTV analysis obtained from ECG monitoring.

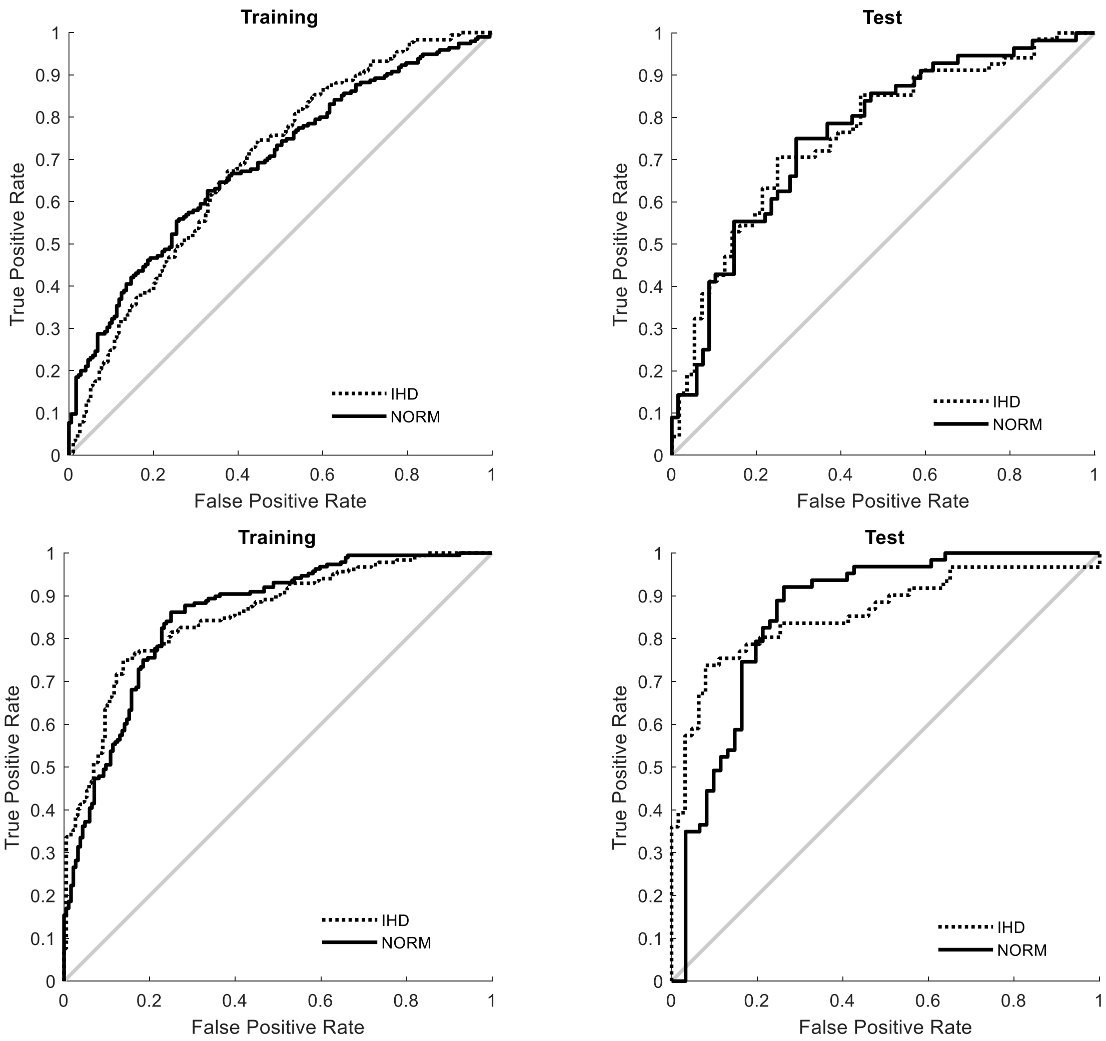

Figure 3 ROC curves in the training and test phases for the ANN2 (top panels) and ANN12 (bottom panels) producing the best accuracies.

In fact, the examination of heart rate variability represents a non-invasive way to evaluate the cardiac function and its assessment has been shown to aid in the clinical diagnosis [4-6]. In particular, it has become an important diagnostic tool useful in early detection of arrhythmias related to cardiovascular disease such as IHD. Patients suffering from IHD present altered circadian rhythm of HRV parameters than normal subjects suggesting an increased sympathetic and/or decreased parasympathetic tone [4-6]. On the other hand, since IHD patients generally present a large number of ectopic beats, usually not considered in the conventional HRV analysis, in this work we would consider the total variability of HR including in the analysis also the supplementary variability introduced by the possible presence of ectopic beats. Thus our results showed a greater variability in IHD than in normal subjects with significantly greater values of RMSSD, HF, HFn and SD1 and lower values of LFn and LF/HF, partially in contrast with the results reported in the literature [4,5] obtained by considering only normal beats. The presence of ectopic beats could explain this large variability as well as the presence of slightly higher (n.s.) values of LF in IHD since these beats, in RR series, often manifest as a derivative spike (short RR followed by compensatory pause), leading to a flat power spectral density contribution at all frequencies (except the zero frequency). Moreover, since LF band is narrower than HF band, the presence of ectopic beats will increase HF power much more than LF power. Such a greater variability is confirmed by further new parameters examined in our work both in time domain, like greater SDRR, RR50 and pRR50 values, and in the non-linear domain like greater SD1/SD2, fractal dimension and lower beta exponent values in IHD patients. The latter was calculated on the whole band (0.002-0.45Hz), contrarily to other authors [6] using only the very low frequency band (0.0001-0.03Hz) in which this parameter showed greater values in IHD patients. Our results also confirmed the outcomes of [5] presenting higher meanRR values in IHD than in normal subjects.

Recently, decision support systems using machine learning techniques have been proposed to identify beats, arrhythmias and cardiovascular diseases [8-23]. In order to identify early subjects suffering from IHD, some DSS based on clinical features, mainly obtained in invasive way [22,23], and/or on HRV parameters [14,20,21] have been proposed for an automatic classification of this disease. In both cases, high accuracy values till 92% in the former [22,23] or 96% in [14] using an adaptive neuro-fuzzy network and 89.5% in [20,21] using ANN techniques were achieved. However, authors in [22,23] although examined a large number of cases, used invasive measurements while [14] and [20] utilized HRV parameters extracted from MIT-BIH arrhythmia database, in which patients with IHD were considered together with dilated cardiomyopathy ones. Moreover, all these latter studies utilized only non-linear HRV features and although they were calculated in several RR tracts, these recordings belonged to few patients (47 patients with nine different cardiac diseases including ischemic/dilated cardiomyopathy). Finally, only 10 IHD patients were considered in [21].

In our study, we would identify which parameters extracted in a non-invasive way from an ambulatory ECG (in which both normal and ectopic RR segments were considered), averaged on 24h and applied to ANN could achieve the highest performance to early distinguish IHD from normal subjects by supporting the result with a large cohort of patients. Furthermore, we would analyze the influence of a known, widely used, clinical non-invasive parameter, LVEF, on ANN performances.

By using as inputs some linear, non-linear or a combination of these HRTV parameters together with age and gender, at first we evaluated six different ANN structures. The best accuracy (71.8 %) and AUC (71.5 %) were obtained excluding among the linear parameters that were significantly correlated for more than 90% (ANN2). This accuracy was lower than that obtained by other authors [20,21] applying ANN to HRV parameters on samples with a very limited number of subjects.

Successively, adding as further input the LVEF feature, a predictive clinical parameter of IHD disease evaluated in non-invasive way [1,3], the classification performance of the examined ANNs increases of about 8% for accuracy and of 15% for AUC. The ANN presenting the highest accuracy (79.8%) and AUC (86.0%) used as input five mixed linear and non-linear HRTV parameters (meanRR, LF, LF/HF, Beta exponent, and SD2) together with age, gender and LVEF, highlighting the features that can be utilized for a high prognostic identification of IHD disease. This performance was comparable to the highest values reported in the literature by Dua et al. [21]. However, it should be underlined that [21] examined only single segments belonging to 10 IHD and 10 healthy subjects, developing an artificial neural network after the application of principal component analysis to only non-linear features extracted from HRV. Our results confirmed the discrimination ability and the validity of ANN classifiers applied to mixed linear/non-linear HRTV parameters on a large set of subjects. Moreover, for the first time, a system able to identify with a low error rate IHD subjects rather than single beats or segments affected by this disease was realized.

Finally, the limited variability of performances, as showed by histograms as the training and test sets changed, highlighted the robustness of the identified ANN, useful for an early identification of IHD patients based only on a few non-invasive parameters.

5. Conclusion

This work confirmed the existence of significant differences between the values of HRTV parameters calculated in normal and in IHD subjects, extending the result to other parameters than those already present in the literature.

IHD patients confirmed the presence of a greater HR variability. Moreover, the study proved the power of machine learning techniques, specifically ANNs, for the classification of IHD patients using non-invasive parameters. Although both linear and non-linear parameters have shown that can be used for the early identification of IHD, the HRTV parameters producing the best performance were a combination of them (meanRR, LF, LF/HF, Beta exponent and SD2). The presence of linear parameters as inputs in both the ANNs producing the best performances, supports the prognostic relevance of this type of features in the specific cardiac disease identification, both when LVEF is used and not. However, the introduction of LVEF allows a significant improvement in ANNs performances providing additional information than that contained in the parameters extracted from HRTV.

Acknowledgment

Work partially supported by Master in Clinical Engineering, University of Trieste, Italy.

References

[1] Kasper DL, Fauci A S, Hauser SL, Longo DL, Jameson JL (2015) Harrison's principles of internal medicine, 19th edn. New York: McGraw-Hill Educational.

[2] Goldenberg I, Goldkorn R, Shlomo N, Einhorn M, Kuperstein R, Klempfner R, Johnson B (2019) Heart rate variability for risk assessment of myocardial ischemia in patients without known coronary artery disease: the HRV-DETECT (Heart Rate Variability for the Detection of Myocardial Ischemia) Study. J Am Heart Assoc 8:1-9.

[3] Knuuti J (2019) 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. The task force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology. Europ Heart J 00:1-71.

[4] Acharya UR, Faust O, Sree V, Swapna G, Martis RJ, Kadri NA, Suri JS (2013) Linear and nonlinear analysis of normal and CAD-affected heart rate signals. Comput Methods Programs Biomed 113:55–68.

[5] Voss A, Schulz S, Schroeder R, Baumert M, Caminal P (2008) Methods derived from nonlinear dynamics for analysing heart rate variability. Phil Trans R Soc 367:277-296.

[6] Guzzetti S, Mezzetti S, Magatelli R, Porta A, De Angelis G, Rovelli G, Malliani A (2000) Linear and non-linear 24 h heart rate variability in chronic heart failure. Auton Neurosc 86:114–119.

[7] Acharya UR, Paul Joseph K, Kannathal N, Lim C-M, Suri JS (2006) Heart rate variability: a review. Med Biol Eng Comput 44:1031–1051.

[8] Christov I, Gómez-Herrero G, Krasteva V, Jekova I, Gotchev A, Egiazarian K (2006) Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. Med Eng Phys 28:876–887.

[9] Rodríguez-Sotelo JL, Castellanos-Domínguez G, Acosta-Medina CD (2012) Recognition of Cardiac Arrhythmia by Means of Beat Clustering on ECG-Holter RecordingsAdvances in Electrocardiograms—Methods and Analysis. PhD. Richard Millis.

[10] Llamedo M, Martinez JP (2011) Heartbeat classification using feature selection driven by database generalization criteria. Trans Biomed Eng 5:616–625.

[11] Jekova I, Bortolan G, Christov I (2004) Pattern recognition and optimal parameter selection in premature ventricular contraction classification. Comput Cardiol 31:357–360.

[12] Behadada O, Chikh M (2013) An interpretable classifier for detection of cardiac arrhythmias by using the fuzzy decision tree. Artif Intel Res 2:45–58.

[13] Wieben O, Afonso V, Tompkins W (1999) Classification of premature ventricular complexes using filter bank features, induction of decision trees and a fuzzy rule-based system. Med Biol Eng Comput 37:560–565.

[14] Kannathal N, Sadasivan KP, Lim CM, Acharya RU, Laxminarayan S (2005) Cardiac state diagnosis using adaptive neuro-fuzzy technique. In: Proceedings IEEE Engineering in Medicine and Biology pp. 3864-3867.

[15] Lin Y, Yang J (2007) ECG Patterns Recognition using Multilayer Perceptron and Classification Tree. In: Proceedings of International Medical Informatics Symposium pp.109–114.

[16] Park J, Kang K (2014) PcHD: Personalized classification of heartbeat types using a decision tree. Comput Biol Med 54:79–88.

[17] Khazaee A, Ebrahimzadeh A (2010) Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features. Biomed Signal Proces 5:252–263.

[18] Acir N (2006) A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems. Expert Syst App 31:150–158.

[19] Jiang W, Kong SG (2007) Block-based neural networks for personalized ECG signal classification. Trans Neural Netw 18:1750–1761.

[20] Acharya RU, Kumar A, Bhat PS (2004) Classification of cardiac abnormalities using heart rate signals. Med Biol Eng Comput 42: 288-293.

[21] Dua S, Du X, Sree V, Ahamed T (2012) Novel classification of coronary artery disease using heart rate variability analysis. J Mech in Med and Biol 12: 1240017-1240036.

[22] Rajeswari K, Vaithiyanathan V, Neelakantan TR (2012) Feature selection in ischemic heart disease identification using feed forward neural networks. Procedia Engin 41:1818-1823.

[23] Kukar M, Kononenko I, Groselj C (1999) Analysing and improving the diagnosis of ischaemic heart disease with machine learning. Art Intell Med 16:25-50.

[24] Fornasa E, Accardo A, Cinquetti M, Merlo M, Sinagra G (2014) HRV Spectral and Fractal Analysis in Heart Failure Patients with Different Aetiologies. Comput Cardiol 41:421-424.

[25] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996) Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 93: 1043-1065.

[26] Woo MA, Stevenson WG, Moser DK, Trelease RB, Harper RM (1992) Patterns of beat-to-beat heart rate variability in advanced heart failure. Am Heart J 123: 704-710.

[27] Higuchi T (1988) Approach to an irregular time series on the basis of the fractal theory. Physica D 31:277-283.

[28] Cusenza M, Accardo A, D'Addio G, Corbi G (2010) Relationship between fractal dimension and power-law exponent of heart rate variability in normal and heart failure subjects. Comput Cardiol 37:935-938.

[29] Haykin S (1995) Neural networks a comprehensive foundation New York:MacMillan College Publishing Company.

[30] Goh AT (1999) Back-propagation neural networks for modeling complex systems. Artif Intel Engin 9:143-151.

[31] Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 4:627–635.