Feature Set Extension for Heart Rate Variability Analysis by Using Non-linear, Statistical and Geometric Measures

Alan Jovic, Nikola Bogunovic
Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3*, HR-10000 Zagreb, Croatia, alan.jovic@fer.hr and nikola.bogunovic@fer.hr*

Abstract. *The goal of this paper is to evaluate the application of a combination of heart rate variability features on successful classification of known heart disorders. We propose an extension over our previous work, which employs 11 features, both from non-linear and linear analysis of heart rate variability. The features were extracted from electrocardiogram recordings and analyzed in Weka system for data mining using several well-known classification algorithms: C4.5 decision tree, Bayesian network, random forest, and RIPPER rules. Significance of each feature is analyzed and the algorithms' success rates are compared. The selected combination of features has a high classification potential.*

Keywords. non-linear analysis; geometric features; ECG classification; classification algorithms, random forest, RIPPER

1. Introduction

The observation of the decrease in heart rate variability (HRV) in patients is often a predictor of coronary problems. It has been associated with both old age and physiological heart problems and it is a popular and efficient method for patient health monitoring [13]. Non-linearity of the heart rhythm is a good descriptor of a patient's overall health and it is this rhythmic pattern that we study in this work. A pioneer investigation performed by [6] introduced the concept of non-linear dynamics into the field of cardiology. Healthy physiological systems have fractal complexity whereas unhealthy biological systems lack the complex non-linear properties and are marked by periodical dynamics and loss of the adaptive capabilities.

In our work, we propose an approach to ECG analysis of HRV by means of several standard statistical HRV measures and several non-linear features which are used to quantify the underlying non-linear processes. This work is an extension of our previous work [7,8] in which we prepared ground for a more thorough investigation of the possibilities of using chaos theory in the successful classification of patient ECG records. It is in this work that we explore a novel combination of features that borrow some of the best qualities from both linear and non-linear analysis. The goals of this work are to:

1. Demonstrate the classification capabilities of some of the HRV statistical, geometric and non-linear features over several types of patient ECG records.
2. Discuss the features' significance and to compare the efficiencies of data mining algorithms on this problem.

In section 2, we present a short overview of the existing work in HRV analysis. Methodology of our work is given in section 3 and the results are presented and discussed in section 4. In section 5 we give a conclusion.

2. Background

The use of HRV analysis for estimating the status of the autonomic system in the heart is well established in literature. The work done by [1] clearly showed the existence of non-linear components in HRV. Today, most of the research done in the field of ECG analysis, particularly classification problems, includes the HRV analysis with at least some non-linear features, because the traditional HRV features do not contain any information about time-directed dynamics [1]. These traditional features have been reviewed and presented in comprehensive guidelines for heart rate variability [13].

Standard time domain features can be roughly divided into statistical and geometric. Geometric features are based on sample density distributions of R-R intervals. R-R intervals are time intervals between two consecutive R peaks in ECG signal. HRV can also be analyzed in frequency domain, wherein usually spectral power is measured. However, the existing guidelines do not evaluate non-linear methods for HRV analysis because this area of ECG analysis is still being researched. Approaches to the evaluation of HRV using non-linear features vary both in the features that are extracted, as well as in the methods used for analysis. It is also a fact that HRV analysis does not contain enough information in order to discern between some specific cardiac dysfunctions, e.g. left bundle branch block and right bundle branch block. Hence, most solutions to successful classification contain at least some additional method applied to the whole ECG signal, as opposed to using only HRV features.

Still, there were many studies conducted that showed the possibility of efficient classification of ECG records based on HRV information alone. Non-linear features that were extracted in these studies include various kinds of entropy measures: approximate entropy, sample entropy, multiscale entropy, and spectral entropy [4].

Also, some of the more readily studied features include correlation dimension D_2, Lyapunov exponents, spatial filling index, and central tendency measure (CTM) [12].

For the purpose of HRV analysis, researchers usually take advantage of standard, internet databases that contain various types of ECG records. The collection of many ECG databases can be accessed from the PhysioBank website [10].

3. Methods

The first part of analysis was to obtain relevant ECG HRV records. For this purpose, we chose several databases from the PhysioBank collection of databases. A total of 100 patient annotation records were prepared as an input to the feature extraction process. The complete list is given in Table 1. This is an extension to our previous work [7], where we analyzed a smaller number of records. Four types of heart conditions were inspected: normal heart rhythm, any arrhythmia, supraventricular arrhythmia and congestive heart failure. Annotation records contain the annotations that were made either by a doctor's inspection or by a machine feature recognition algorithm. The most common annotation is the time of an R peak in the ECG signal. R-R intervals were extracted from the designated times by using ECE platform [8].

For each ECG record, first 5 minutes of the R-R interval time series were analyzed. Depending on the heart rate, this constitutes to around 500 R-R intervals. The records were not searched for time segments with anomalous rhythm. Thus, the 5 minute intervals do not necessarily contain the anomalous rhythm. This fact poses a significant challenge and limitation to the classifier. However, it is our opinion that in a real hospital environment it is necessary to have a reliable prognosis for a patient disorder, even if the disorder itself is not momentarily present in the ECG signal.

It is the recommendation of [13] to use four HRV features, three statistical measures and one geometric measure, in any serious HRV analysis. This was respected and pursued in this work. ECG Chaos Extractor platform implements extraction algorithms for four heart rate variability features and seven non-linear features. Standard heart rate variability features that were extracted include: SDNN, RMSSD, pNN20, and

Table 1. Patient records used in the study.

Heart condition	Internet database	Annotation records
Normal heart rhythm	MIT-BIH Normal Sinus Rhythm Database, Normal Sinus Rhythm RR Interval Database	16265, 16272, 16273, 16420, 16483, 16539, 16773, 16786, 16795, 17052, 17453,17693, 18177, 18184, 19088, 19090, 19093, 19094, 19140, 19830, nsr001, nsr002, nsr003, nsr004, nsr005
Arrhythmia	MIT-BIH Arrhythmia Database	100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 111, 112, 113, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 200, 201
Supraventricular arrhythmia	MIT-BIH Supraventricular Arrhythmia Database	800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 840, 841
Congestive heart failure	BIDMC Congestive Heart Failure Database, Congestive Heart Failure RR Interval Database	chf1, chf2, chf3, chf4, chf5, chf6, chf7, chf8, chf9, chf10, chf11, chf12, chf201, chf202, chf203, chf204, chf205, chf206, chf207, chf208, chf209, chf210, chf211, chf212, chf214

HRV triangular index (HTI). Non-linear features are: spatial filling index (SFI), central tendency measure (CTM), correlation dimension D_2, and four features for approximate entropy (ApEn1-ApEn4).

For non-linear methods in HRV analysis, we use phase space reconstruction. Description of the algorithms for non-linear features analyzed in this work is given in [8]. We give here only the explanation of the additional features that were extracted: SDNN, RMSSD, pNN20 and HRV triangular index. They are elaborated in subsection 3.1. Classification algorithms are briefly explained in subsection 3.2.

3.1. Statistical and geometric measures

SDNN (Standard Deviation of the NN interval) is the simplest statistical HRV feature to calculate. NN stands for time interval between consecutive normal sinus heart beats. In most heart rhythms, NN interval is equivalent to the R-R interval. The standard deviation reflects all the cyclic components responsible for variability in the period of recording. It can be calculated for 24 hours long-term recordings and for short-term, five minutes recordings [13]. In our case, we used it to calculate the standard deviation of the short-term (five minutes) recording of R-R intervals.

RMSSD is the square root of the mean squared differences of N successive NN (or R-R) intervals. Thus, it can be calculated using the expression:

$$RMSSD = \frac{\sqrt{\sum_{i=1}^{N-1}(x_{i+1} - x_i)^2}}{N} \qquad (1),$$

where x_i denotes the length of an R-R interval with index i.

pNN20 is the ratio of the number of interval differences of successive NN (or R-R) intervals that are greater than 20 ms and the total number of NN (or R-R) intervals. In the work [9], it is concluded that the enhanced discrimination between a variety of normal and pathological conditions is obtained by using pNN thresholds as low as 20 ms or less.

HRV triangular index (HTI) is a geometric measure of HRV. It measures the integral of the density distribution. When approximated, it calculates the number of all NN (R-R) intervals divided with the total number of NN intervals in the modal bin of the histogram. Obviously, HRV triangular index is dependent on the length of the bin, i.e. sampling frequency of the ECG record. If the bin size is different from the most common sampling frequency, it should be noted [13].

It was recommended that in order to have a representative HRV triangular index, at least a 20 minutes long R-R interval recording should be analyzed. This was not pursued in our work due to an increased demand on the extraction time and because other features do not have such recommendations.

3.2. Classification algorithms

With the advantage of using the Weka system, we were in position to examine the possibilities of successful classification of records by using many classification algorithms. In the end, it was decided that the following four algorithms should be used: C4.5 decision tree, Bayesian network classifier, random forest (RF) and Repeated Incremental Pruning to Produce Error Reduction (RIPPER).

C4.5 (J48 in Weka) is the landmark decision tree algorithm developed by [11]. C4.5 was used with error based pruning (default) and a minimum amount of five instances per leaf. Five instances per leaf are used instead of the standard two in order to ensure that only relevant leaves are taken into consideration. This improved stability and reliability of the method, but reduced the classification accuracy by a small degree.

Bayesian network is a known probabilistic graphical model classifier based on the Bayesian theorem and its implications [5]. The network is constructed using several parameters, including the type of estimator and search method. Simple estimator based on maximum likelihood and hill climbing search methods were employed.

Random forest is a state-of-the-art classifier developed by [2]. It is composed of a number of decision trees that choose their splitting attributes from a random subset of attributes at each internal node. The best split is taken among these randomly chosen attributes and the trees are built without pruning. Based on the bootstrap sampling procedure, random forest ensures at the same time the smallest obtainable bias and very low data variance which gives them optimal classification results. They are widely used in various classification problems, especially in domains with large numbers of attributes and instances. The only negative side of random forest is that its model is not clearly any single tree. It is a majority vote from the whole ensemble, thus virtually disabling any explicit rule-based interpretation (although particular trees in the forest can still be analyzed). We include random forest in our work for the purpose of exploring the upper limits of classification accuracy for our feature set.

RIPPER [3] is a rule-based classifier intended for successful separate-and-conquer based strategy over the feature space. It is an algorithm with excellent results over a variety of representative datasets. The main advantages of using this procedure are a clear set of classification rules and its speed. Its accuracy is in every way comparable to the landmark decision tree algorithm C4.5. RIPPER was used with default parameters in Weka system.

4. Experimental results

The output results from Weka system include standard statistical measures. We used total classification accuracy for four patient classes and total classification accuracy (positive prediction), sensitivity and specificity for two patient classes' case. Several analyses were performed.

First, classification algorithms on all four patient record types were employed. The number of feature vectors per record type is 125 because five feature vectors per record were used. These five feature vectors contain features for representative time intervals $T = \{1, 2, 5, 10, 20\}$ between two R peaks. The results are given and discussed in subsection 4.1.

Second, we used classification algorithms on all four patient record types with a single value for $T = 1$. We give and discuss the results in section 4.2.

Third, we evaluate the features in two-class classification case, with $T = \{1, 2, 5, 10, 20\}$. Herein, one class of patients has a normal heart rhythm and the other class includes patients with a heart disorder. Analysis is performed on the following combinations: normal-arrhythmia, normal-supraventricular arrhythmia, normal-congestive heart failure. The results are presented in subsection 4.3.

Fourth, we evaluate the extracted features themselves, as elaborated in subsection 4.4.

In all four subsections, for classification purposes, a 10-fold cross-validation technique has been used in order to randomize the input samples and obtain representative classification accuracy.

4.1. Four classes' classification, five intervals

Classification results for four patient classes in the case of a set of five intervals are presented in Table 2. All the methods worked very well and achieved high classification accuracy. C4.5 gave worst results due to the rigorous limitation to the minimum number of instances in the leaves. However, this limitation may be justified, because a stable and reliable classifier is requested, i.e. one that would never overfit. Random forest of only five trees achieved almost perfect classification of instances. Here are some of the rules obtained by RIPPER algorithm which cover most of the samples:

Table 2. Four classes classification results, $T = \{1, 2, 5, 10, 20\}$

Classification algorithm	Classification accuracy, %
C4.5	89.4
Bayesian network	96.2
Random forest (5 trees)	99.8
RIPPER	93.6

1. (HRV_triangular_index >= 20.52381) and (spatial_filling >= 3.30398) and (approximate_entropy1 >= 1.432138) => signal_type=Atrial_arrhythmia (72.0/0.0)
2. (HRV_triangular_index >= 8.781818) and (RMSSD <= 0.068436) and (pNN20 >= 0.082692) and (approximate_entropy3 <= 1.223611) => signal_type=Normal (70.0/0.0)
3. (RMSSD <= 0.075541) and (approximate_entropy4 <= 1.107562) and (approximate_entropy2 >= 1.065076) => signal_type=Congestive_heart_failure (54.0/0.0)
5. => signal_type=Supraventricular_arrhythmia (125.0/0.0)

As we can see, explicit and understandable rules are obtainable from the dataset with high accuracy. The total number of rules for RIPPER in this analysis was 14.

4.2. Four classes' classification, one interval

Classification results for four patient classes in the case of one interval are presented in Table 3. We decided to present results for $T = 1$

because we established no significant difference in classification results with the choice of interval. The results from Table 3 have clearly much lower accuracy than in the case of a set of intervals given in Table 2.

Table 3. Four classes classification results, *T* = 1

Classification algorithm	Classification accuracy, %
C4.5	56
Bayesian network	58
Random forest (40 trees)	69
RIPPER	52

It is also obvious that the number of learning samples (instances) is five times lower, thus reducing the classification power of all the algorithms. The results were still in the range far above random choice (25%), with random forest displaying the best fitting. Only four rules were constructed with RIPPER algorithm and these do not separate samples acceptably.

4.3. Two classes' classification

Classification results for two patient classes in the case of a set of five intervals are presented in Table 4. Results in the two classes' case show an increase in classification accuracy compared to the four classes' case, as expected. General observation is that the best results are achieved when discerning between normal heart rhythm and arrhythmia. Nevertheless, the results for the other two record pairs follow close behind, all of them showing classification accuracy higher than 90%. As an example, a set of rules that give perfect classification in the case of normal heart rhythm and arrhythmia is given here:

1. (HRV_triangular_index <= 20.421053) and (spatial_filling <= 7.682874) => signal_type=Normal (100.0/0.0)
2. (HRV_triangular_index <= 12.615385) => signal_type=Normal (20.0/0.0)
3. (approximate_entropy1 <= 1.290091) => signal_type=Normal (5.0/0.0)
4. => signal_type=Atrial_arrhythmia (125.0/0.0)

5.4. Feature evaluation

It is important to determine which of the extracted 11 features contribute in the greatest extent to the quality of the obtained results. Because of the difficulty in establishing objective contribution of a particular feature to the classification accuracy, we opted for a rather simple, but justifiable approach. As a test bench, we use feature vectors obtained from four classes of patients. We analyzed classification results when the features were allowed only in pairs. Hence, 55 feature pairs were formed out of the 11 features and their classification accuracy was calculated.

Algorithm used in the classification procedure was C4.5, the same one that was employed in previous analyses. In Table 5, 12 feature pairs that had total classification accuracy of 50% and above are presented. We can conclude from Table 5 that HTI appears in most of the pairs with the best classification accuracy, thus making this feature the optimal choice for the analysis. Other two features that we observed to have demonstrated good classification capabilities were pNN20 and D_2.

Table 4. Two classes classification results *T* = {1, 2, 5, 10, 20}

Records	Algorithm	Total classification accuracy, %	Sensitivity, %	Specificity, %
Normal and Arrhythmia	C4.5	91.2	91.9	90.6
	Bayesian network	98.8	97.6	100.0
	Random forest (3 trees)	100.0	100.0	100.0
	RIPPER	100.0	100.0	100.0
Normal and Supraventricular arrhythmia	C4.5	96.0	97.5	94.6
	Bayesian network	92.4	91.3	92.7
	Random forest (9 trees)	100.0	100.0	100.0
	RIPPER	97.6	99.2	96.1
Normal and Congestive heart failure	C4.5	95.6	93.2	98.3
	Bayesian network	95.6	92.5	99.1
	Random forest (20 trees)	100.0	100.0	100.0
	RIPPER	96.4	96.0	96.8

Table 5. Top 12 ranking feature pairs

Pair number	Feature pairs	Classification accuracy, %
10	SFI, HTI	58
19	D_2, HTI	55
34	ApEn1, HTI	55
40	ApEn2, HTI	55
45	ApEn3, HTI	55
49	ApEn4, HTI	55
54	RMSSD, HTI	55
27	CTM, HTI	54
52	SDNN, HTI	54
55	pNN20, HTI	52
33	ApEn1, pNN20	50
39	ApEn2, pNN20	50

5. Conclusion

We have demonstrated the efficiency of using the combination of 11 features taken from HRV statistical, geometric and non-linear measures. Features that were established to contribute the most to successful classification were: HRV triangular index, correlation dimension D_2 and pNN20. Among the employed classifiers, the highest score was obtained from random forest classifier, which was expected. Nevertheless, all the algorithms gave high classification results, especially in the case of two classes of records, where the results were high above 90%.

In the case where there was not enough learning material (only one interval analyzed), the classification results were mostly not satisfactory. This suggests that a combination of several intervals should be analyzed at the same time to obtain favorable results. Further work should be directed towards introduction of non-linear features combinations in a real patient monitoring system.

6. References

[1] Braun C, Kowallik P, Freking A, Hadeler D, Kniffki KD, Meesmann M. Demonstration of Nonlinear Components in Heart Rate Variability of Healthy Persons. Am J Physiol Heart Circ Physiol 1998; 275(5):1577-1584.

[2] Breiman L. Random forests. Machine Learning 2001; 45:5-32.

[3] Cohen WW. Fast effective rule induction. In: Prieditis A, Russell SJ, editors. Proceedings of the Twelfth International Conference on Machine Learning; 1995 Jul 9-12; Tahoe City, California, USA. Morgan Kaufmann; 1995. p. 115-123.

[4] Costa M, Goldberger AL, Peng CK. Multiscale Entropy Analysis of Complex Physiologic Time Series. Phys Rev Lett 2002; 89 (6):068102.

[5] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. Machine Learning 1997; 29(2-3):131-163.

[6] Goldberger AL. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. Lancet 1996; 11:1312-1314.

[7] Jovic A, Bogunovic N. Analysis of ECG Records using ECG Chaos Extractor Platform and Weka System. In: Luzar-Stiffler V, Hljuz Dobrić V, editors. Proceedings of the 30th International Conference on Information Technology Interfaces; 2008 Jun 23-26; Cavtat, Croatia. Zagreb: SRCE University Computing Centre, University of Zagreb; 2008. p. 347-352.

[8] Jovic A, Bogunovic N. Feature Extraction for ECG Time-Series Mining Based on Chaos Theory. In: Luzar-Stiffler V, Hljuz Dobrić V, Bekić Z, editors. Proceedings of the 29th International Conference on Information Technology Interfaces; 2007 Jun 25-29; Cavtat, Croatia. Zagreb: SRCE University Computing Centre, University of Zagreb; 2007. p. 63-68.

[9] Mietus JE, Peng CK, Henry I, Goldsmith RL, Goldberger AL. The pNNx files: re-examining a widely used heart rate variability measure. Heart 2002; 88:378-380.

[10] PhysioBank: Physiologic signal archives for biomedical research. http://www.physionet.org [15/01/2009]

[11] Quinlan JR. C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers; 1993.

[12] Seker R, Saliu S, Birand A, Kudaiberdieva G. Validity Test for a Set of Nonlinear Measures for Short Data Length with Reference to Short-Term Heart Rate Variability Signal. Journal of Systems Integration 2000; 10:41-53.

[13] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology. Heart rate variability guidelines: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal 1996; 17:354-381.