Risk Stratification for Arrhythmic Sudden Cardiac Death in Heart Failure Patients using Machine Learning Techniques

George Manis[1], Stavros Nikolopoulos[2], Petros Arsenos[3], Konstantinos Gatzoulis[3], Polychronis Dilaveris[3], Christodoulos Stefanadis[3]

[1] Dept. of Computer Science and Engineering, University of Ioannina, Greece
[2] School of Electrical and Computer Eng., National Technical University of Athens, Greece
[3] First Dept. of Cardiology, National Kapodistrian University of Athens, Greece

Abstract

Arrhythmic Sudden Cardiac Death (SCD) is still a major clinical challenge even though much research has been done in the field. Machine learning techniques give a powerful tool for stratifying arrhythmic risk. We analyzed 40 Holter recordings from heart failure patients, 20 of which were characterized as high arrhythmia risk after 16 months follow up. The two groups (high and low risk) were not statistically different in basic clinical characteristics. We performed windowed analysis and computed 25 Heart Rate Variability (HRV) indices. We fed these indices as input to two classifiers: Support Vector Machines (SVM) and Random Forests (RF). The classification results showed that the automatic classification of the two groups of subjects is possible.

1. Introduction

Today, heart problems have become the most frequent cause of death in the western world. The number of sudden cardiac deaths in the United States is estimated between 200.000 and 500.000 per year, with the 50% to 70% being due to mechanisms related to arrhythmias [1]. Even though much research has been done in the field, risk stratification of arrhythmic SCD is a major clinical challenge [2].

The research in the field of heart rate variability has been very active the past decades. Today the research in the field is still active, with many new interesting papers appearing every day. For the analysis of HRV, a large number of HRV measure has been proposed and used. Almost two decades ago, a review on heart rate variability summarizing the knowledge in the field and suggesting guidelines for HRV analysis was published [3]. Since then many other metrics has been proposed and reviews in the field have been published including [4, 5].

The crucial question *which HRV measure to use* has not been answered yet and we do not think it is possible to be answered, at least in the near future. Each measure extracts different information and all this information must be combined and summarized. However, this is an especially difficult task, since the information can be sometimes huge, but most important we are not yet in the position to describe the rules and propose procedures.

One solution to the problem, the most realistic up to date, is to use machine learning techniques in order to automatically combine the information extracted from the HRV methods so that inherent and not well described relations hidden in the available information to be discovered and exploited.

In this paper we study the arrhythmic sudden cardiac death in heart failure patients using machine learning techniques. We analyzed 40 Holter recordings from heart failure patients. The half of them (20 subjects) were characterized as *high arrhythmia risk* after 16 months of follow up. The rest (20 subjects) were characterized as *low arrhythmia risk*. The two groups were not statistically different in basic clinical characteristics.

We performed windowed analysis and computed 25 heart rate variability indices. We used these data as input to a support vector machine and a random forest classifier. The two classifiers achieved a high level of accuracy showing that the automatic classification of high and low risk patients is possible.

2. Classification methods

2.1. Support vector machines

If two sets can be linearly separated there exists a linear function of the form:

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \quad (1)$$

such that $y_i f(\mathbf{x}_i) \geq 0$, or $f(\mathbf{x}_i) \geq 0$ for $y_i = +1$ and $f(\mathbf{x}_i) < 0$ for $y_i = -1$. Vector $\mathbf{w}$ and scalar b represent

the hyperplane $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$ separating the two classes.

Since there are more than one hyperplane which can separate the two classes, the SVM classifier selects the hyperplane which maximizes the separating margins [6,7] by minimizing the cost function:

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}\|\mathbf{w}\|^2 \quad (2)$$

subject to the separability constraints

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad i = 1, ..., l. \quad (3)$$

If the training data is not completely separable by a hyperplane a set of slack variables $\xi_i \geq 0$, $i = 1, ..., l$ is introduced which represent the amount by which the linearity constraint is violated:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, ..., l. \quad (4)$$

In that case, the cost function is modified to take into account the extent of the constraint violations. Hence, the function to be minimized becomes:

$$J(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \quad (5)$$

subject to the constraints in (4).

Sometimes it is unlikely that a hyperplane will yield a good classifier and we need a decision boundary with more complex geometry. We can do this by mapping the attribute vector into a new space of higher dimensionality and look for a hyperplane in that new space. This is what we call a *kernel-based* SVM [8, 9]. Let $\Phi(\cdot)$ be a non linear operator mapping the input vector $\mathbf{x}$ to a higher dimensional space. The optimization problem for the new points $\Phi(\mathbf{x})$ becomes:

$$min \;\; J(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \quad (6)$$

subject to the constraints:

$$y_i(\mathbf{w}^T\Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, ..., l. \quad (7)$$

A widely used non-linear kernel is the *Radial Basis Function (RBF)*:

$$k(\mathbf{x_i}, \mathbf{x_j}) = exp(-\gamma||\mathbf{x_i} - \mathbf{x_j}||^2) \quad (8)$$

An SVM which uses an RBF kernel has two parameters which should be estimated: the parameter C of equation 6 and the parameter γ from equation 8.

2.2. Random forests

A *random forest* is a classifier which consists of many decision trees. Each tree is constructed using a subset of samples selected from the dataset. The tree is built to the maximum size without pruning. A subset of the set of the features is used as candidate splitters in the nodes of the tree. Splitting is done according to a feature ranking mechanism (e.g. Gini Index). The samples that were not selected for the construction of the tree constitute the test set of the tree. Once the forest is constructed samples from the test set are given as input to each one of the trees in the forest. Each tree classifies the sample and actually *votes* for the class in which the sample belongs. The number of votes for each class decides the class in which this sample belongs [10,11]. A random forest has one parameter which should be estimated: the number of trees which constitute the random forest.

3. Description of data

We study 40 patients with ischemic heart failure. All recordings were acquired in the First Department of Cardiology, Medical School, National and Kapodistrian University of Athens, supervised by clinical experts. Patients underwent physical examination, chest X Ray, blood and biochemical tests, 12 lead-ECG, ECHO, Signal Averaged ECG (SAECG) and Holter Monitoring (HM), while personal, family history and medications were registered. All patients provided informed consent and the study was approved by our institution's Ethics Committee. From those patients, 20 were characterized as *high risk* for SCD after 16 months of follow up and the rest 20 arrhythmia free patients considered as *low risk*. Further details on the data can be found on table 1.

Table 1. Description of the dataset

	High Risk	Low Risk	p-value
Recordings	20	20	
Age (years)	64.9±13.1	66.0±15.8	0.813
Gender	18m-2f	17m-3f	0.633
LVEF (%)	29.6±9.8	34.5±9.3	0.117
Heart Rate (bpm)	73.2±14.5	66.1±10.3	0.095
Standard QRS (ms)	130±27	122±40	0.491
QT-corrected (ms)	428±35	422±33	0.569

4. HRV analysis methods as features

There is a large number of methods used for HRV analysis. We tried to cover a wide variety of methods in our experiments and not to omit at least any of the most widely

used ones. We also used several non-linear methods, since non-liner methods can sometimes extract more sensitive information. A list of the methods used as attributes in our classification procedure follows, grouped in different categories.

Statistical Methods [3]: the standard deviation of NN intervals ($sdnn$), the root mean square of successive differences ($rmssd$), the probability the difference of successive intervals be more than 50ms ($pNN50$) and the standard deviation of successive differences ($sdsd$), as well as the moments m_1,m_2,m_3 and m_4.

Geometrical Methods [3]: the triangular index defined as the integral of the density distribution divided by the maximum of the density distribution (ti).

Frequency Analysis Methods [3]: power in low frequency range (LF_n) and power in high frequency range 0.15-0.4 Hz (HF_n) both in normalized units.

Non-linear Methods: approximate entropy (ApEn) [12] modified to avoid self comparisons as suggested in [13] computed for $m = 2$ and $r = 0.2$, a_1 and a_2 from detrented fluctuation analysis (DFA) [14], the β exponent [15]. We also computed sd_1 and sd_2 from the 2d $Poincaré$ plot and v_{max} and v_{spread} from the 3d $Poincaré$ plot. The metric v_{max} is extracted by constructing a 3-dimensional plot, in which the azimuthal plane corresponds to the normal 2-dimensional $Poincaré$ plot. We divide this plot into subsquares and then, we estimate the cardinal number of points in each subsquare. We plot these cardinal number values in the z-axis. V_{max} corresponds to the maximum cardinal number and v_{spread} to the width of the plot. We also computed Hurst exponent with the RS method and the wavelets method, the decorrelation time and the mobility and complexity as defined by Hjorth [16].

Table 2. Classification using SVM

Features	25		
Selected	sd_1,sd_2,LF_n,m_2,$sdsd$,$pNN50$		
Parameters	RBF Kernel, C=125, γ=10		
Validation	10–fold cross validation		
	Low Risk	High Risk	Average(weighted)
TP rate	0.9	0.85	0.875
FP rate	0.15	0.1	0.125
Precision	0.857	0.895	0.876
Fmeasure	0.878	0.872	0.875
under ROC	0.875	0.875	0.875
Accuracy	87.5%		

Table 3. Classification using RF

Features	25		
Selected	sd_1,sd_2,LF_n,m_2,$sdsd$,$pNN50$		
Parameters	20 trees, 5 random features per node		
Validation	10–fold cross validation		
	Low Risk	High Risk	Average(weighted)
TP rate	0.9	0.8	0.85
FP rate	0.2	0.1	0.15
Precision	0.818	0.889	0.854
Fmeasure	0.857	0.842	0.85
under ROC	0.87	0.87	0.87
Accuracy	85%		

5. Classification results

For our experiments the Weka [17] software was used. We performed experiments with SVMs and RFs. We divided each signal into a number of non-overlapping windows. For each one of these windows we computed all 25 metrics described in section 4. Thus from each signal $\lfloor N/w \rfloor$ vectors of 25 features were produced, where N is the length of the signal and w the size of the window.

We selected a window size of $w = 1024$, which is approximately 15-20 minutes of recording, a time period which is not far away from the time period in which the heart beat can be considered stationary, but long enough for methods which need longer signals to extract reliable information. For each feature we selected the most often value for unimodal distributions or the average of the modes when the distribution is not unimodal (*mode statistics*).

From the set of 25 features we selected the most important ones using a *Chi Squared attribute evaluator*, evaluating the worth of an attribute by computing the value of the chi-squared statistic with respect to the class.

Results are summarized on tables 2 and 3 respectively. The features that were selected and participated in the classification procedure were: sd_1, sd_2, LF_n, m_2, $sdsd$, $pNN50$. For the SVM a RBF function with $C = 125$ and $\gamma = 10$ was used. Each RF was consisted of 25 trees, with 5 features per node, randomly selected. Ten-fold cross validation was used to obtain the classification performance.

In these tables we can see several metrics used to evaluate a classification experiment. *Accuracy* is the percentage of the correctly classified samples. The *True Positive (TP)* index or *Recall* or *Sensitivity* is the ratio of the number of samples classified in a class to the total number of samples which truly belong in this class. The *False Possitive (FP)* or *Specificity* index is the ratio of the number of samples classified in a class but do not belong in this class to the

total number of samples which do not belong in this class. The *Precision* index is the ratio of the number of samples which trully belong in a class to the number of samples classfied in this class. The *Fmeasure* is a combined index computed from *Recall* and *Precission* from the following formula:

$$Fmeasure = \frac{2 \times Recall \times Precision}{Recall + Precission}.$$

Finally the *Area Under ROC Curve* is the area under the TP versus the FP indices as the discrimination threshold varies.

The experimental results show that both classifiers achieved a remarkable classification accuracy. Slightly better accuracy was achieved by the SVM classifier. From the experimental analysis it is obvious that the classification of low and high risk heart failure patients for arrhythmic sudden cardiac death is possible and can be done with remarkable accuracy.

6. Conclusions

We used machine learning techniques to classify heart failure patients in two groups: those with a high risk for arrhythmic sudden cardiac death and those with low risk. As input to the classifiers we gave 25 HRV measures, including all popular as well as many non-linear ones. We performed windowed analysis with non-overlapping windows of 1024 beats and for selecting the value of each feature we used mode statistics. As classifiers we used SVMs and RFs. Both classifiers achieved very good classification performance, with the SVMs to perform slightly better. We can safely conclude that the classification between low and high risk heart failure patients for arrhythmic sudden cardiac death is possible and reliable.

References

[1] Goldberger J, Buxton A, Cain M, Costantini O, Exner D, Knight B, Lloyd-Jones D, Kadish A, Lee B, Moss A, Myerburg R, Olgin J, Passman R, Rosenbaum D, Stevenson W, Zareba W, Zipes D. Risk stratification for arrhythmic sudden cardiac death– identifying the roadblocks. Circulation 2011;123:2423–2430.

[2] Arsenos P, Gatzoulis K, Dilaveris1 P, Manis G, Tsiachris D, Archontakis S, Vouliotis AH, Sideris S, Stefanadis C. Arrhythmic sudden cardiac death: Substrate, mechanisms and current risk stratification strategies for the post-myocardial infarction patient. Hellenic J Cardiol 2013;54:301–315.

[3] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Circulation 1996; 93:1043–1065.

[4] Teich MC, Lowen SB, Jost BM, Vibe-Rheymer K, Heneghan C. Heart Rate Variability: Measures and Models. Addison Wesley Higher Education, 2006.

[5] Acharya UR, Joseph PK, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Engineering and Computing 2006;44(12):1031–1051.

[6] Cortes C, Vapnik VN. Support vector networks. Machine Learning 1995;20:1–25.

[7] Christianini N, Shawe-Taylor J. Support Vector Machines and other kernel-based methods. Cambridge University Press, 2000.

[8] Kondor R, Jebara T. A kernel between sets of vectors. In 20^{th} International Conference on Machine Learning (ICML). Washington DC, USA, 2003; .

[9] Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks 2001;12(2):181–202.

[10] Breiman L. Random forests. Machine Learning 2001;45:5–32.

[11] Tan PN, Steinbach M, Kumar V. Introduction to data mining. Addison Wesley Higher Education, 2006.

[12] Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America 1991;88(6):2297–2301.

[13] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Physiol Heart Circ Physiol 2000;278:2039–2049.

[14] Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 1995;5(1):82–87.

[15] Heneghan C, McDarby. G. Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. Physical Review E 2000;62(5):6103–6110.

[16] Spehr W, Sartorius H, Berglund K, Hjorth B, Kablitz C, Plog U, Wiedemann P, Zapf K. EEG and haemodialysis. a structural survey of EEG spectral analysis, Hjorth's EEG descriptors, blood variables and psychological data. Electroencephalogr Clin Neurophysiol 1977;43(6):787–97.

[17] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor Newsl November 2009;11(1):10–18.

Address for correspondence:

George Manis
Dept. of Computer Science and Engineering
University of Ioannina
P.O. Box 1186, Ioannina, 45110, Greece
manis@cs.uoi.gr