

# Robustness of Support Vector Machine-based Classification of Heart Rate Signals

Argyro Kampouraki, Christophoros Nikou\* and George Manis

University of Ioannina, Department of Computer Science,  
P.O. Box 1186, 45110 Ioannina, Greece,  
Phone: +30 26510 98802, Fax: +30 26510 98890, email: cnikou@cs.uoi.gr.

**Abstract**—In this study, we discuss the use of Support Vector Machine (SVM) learning to classify heart rate signals. Each signal is represented by an attribute vector containing a set of statistical measures for the respective signal. At first, the SVM classifier is trained by data (attribute vectors) with known ground truth. Then, the classifier learnt parameters can be used for the categorization of new signals not belonging to the training set. We have experimented with both real and artificial signals and the SVM classifier performs very well even with signals exhibiting very low signal to noise ratio which is not the case for other standard methods proposed by the literature.

## I. INTRODUCTION

Heart Rate Variability (HRV) analysis is based on measuring the variability of heart rate signals and more specifically, the variability in intervals between R peaks of the electrocardiogram (ECG), referred as RR intervals. Several techniques have been proposed for the investigation of evolution of features of the HRV time series. A survey of statistical methods, based on the estimation of the statistical properties of the beat-to-beat time series, can be found in [1]. These methods describe the average statistical behavior of the signal over a considered time window. Spectral methods [2], based on FFT or standard autoregressive modeling, were also proposed. More recently, nonlinear approaches, including Markov modeling [3], entropy-based metrics [4], [5], the mutual information measure [6] and probabilistic modeling [7], [8] were presented to examine heart rate fluctuations. Other methods include the application of the Karhunen-Loève transformation [9] or modulation analysis [10], [11].

In this study, we investigate the potential benefit of using a support vector machine (SVM) learning [12], [13] to classify heart rate signals. Support vector classifiers are based on recent advances on statistical learning theory [14]. They use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. In the last decade, SVM learning has found a wide range of applications [15], including image segmentation [16] and classification [17], object recognition [18], image fusion [19] and stereo correspondence [20]. Based on our previous work on support vector classification

of heart beat series [21], the evaluation of SVM to the presence of noise is examined here. We also compare SVM classification to the classification obtained by the Learning Vector Quantizer (LVQ) neural network [22].

## II. METHODS

Support vector learning strategy is a principled and very powerful method that has outperformed most other systems in a wide variety of applications [15]. The learning machine is given a training set of examples (or inputs), belonging to two classes, with associated labels (or output values). The examples are in form of attribute vectors and the SVM finds the hyperplane separating the input data and being furthest from both convex hulls. If the data are not linearly separable a set of slack variables is introduced representing the amount by which the linear constrained is violated by each data point.

In this study we are concerned with a two-class pattern classification problem. Let vector  $\mathbf{x} \in \mathbb{R}^n$  denote a pattern to be classified and let scalar  $y$  denote its class ( $y \in \{\pm 1\}$ ). Also let  $\{(\mathbf{x}_i, y_i), i = 1, \dots, l\}$  denote a set of  $l$  training examples. The problem is how to construct a decision function  $f(\mathbf{x})$  that correctly classifies an input pattern that is not necessarily in the training set.

### A. Linear SVM classifiers

If the training patterns are linearly separable 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.

While there may exist many hyperplanes separating the two classes the SVM classifier finds the hyperplane that maximizes the separating margins between the two classes [12], [13]. This hyperplane can be found 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

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

\*Asterisk indicates corresponding author.

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

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

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). Here,  $C$  gives the significance of the constraint violations with respect to the distance between the points and the hyperplane and  $\xi$  is a vector containing the slack variables. The cost function in (5) is called *structural risk* and is a trade-off between the empirical risk (the training errors reflected by the second term) with model complexity (the first term) [23]. The purpose of using model complexity to constrain the optimization of empirical risk is to avoid overfitting, a situation in which the decision boundary corresponds to the training data and thereby fails to perform well on data outside the training set.

The problem in (5) with the constraints in (4) can be solved by introducing Lagrange multipliers. With some manipulation, it can be shown that the vector  $\mathbf{w}$  is formed by linear combination of the training vectors:

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \quad (6)$$

where  $\alpha_i \geq 0, i = 1, \dots, l$  are the Lagrange multipliers associated with the constraints in (4). The Lagrange multipliers are solved for the dual problem of (5), which is expressed as:

$$\max_{\alpha_i} \left\{ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j (y_i y_j \mathbf{x}_i \mathbf{x}_j) \alpha_j \right\} \quad (7)$$

subject to the constraints:

$$\alpha_i \geq 0, \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (8)$$

The cost function to be maximized in (7) is convex and quadratic with respect to the unknown parameters  $\alpha_i$  and in practice, it is solved numerically through quadratic programming.

Note that only a few parameters  $\alpha_i$  will have values satisfying the constraints in (8), that is will be nonzero. The corresponding training vector  $\mathbf{x}_i$  is called *support vector*. Vector  $\mathbf{w}$  is computed from (6) while scalar  $b$  is computed from  $y_i(\mathbf{w} \mathbf{x}_i + b) = 1$  for any support vector. The classification of a vector  $\mathbf{x}$  outside the training set is performed by:

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^l (\alpha_i y_i \mathbf{x} \mathbf{x}_i + b) \right) \quad (9)$$

## B. Kernel-based SVM classifiers

For many datasets, it is unlikely that a hyperplane will yield a good classifier. Instead, we want a decision boundary with more complex geometry. One way to achieve this is to map the attribute vector into some new space of higher dimensionality and look for a hyperplane in that new space, leading to kernel-based SVMs [24], [25]. The interesting point about kernel functions is that although classification is accomplished in a space of higher dimension, any dot product between vectors involved in the optimization process can be implicitly computed in the low dimensional space [13].

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:

$$\min J(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (10)$$

subject to the constraints:

$$y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, l. \quad (11)$$

Following the same principles as in the linear case, we note that the only form in which the mapping appears is in terms of  $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}_j)$ . That is, the mapping appears only implicitly through the kernel function  $K(\cdot, \cdot)$ . There are a variety of possible kernels. However, when choosing a kernel it is necessary to check that it is associated with the inner product of some nonlinear mapping [23]. Some typical choices for kernels are polynomials and radial basis functions.

Finally, the dual problem to be solved is:

$$\max_{\alpha_i} \left\{ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j (y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)) \alpha_j \right\} \quad (12)$$

subject to the constraints:

$$\alpha_i \geq 0, \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (13)$$

and the classifier becomes:

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^l (\alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b) \right) \quad (14)$$

## C. Feature extraction

Standard classification techniques applied to the study of HRV rely on signal statistics. In this study we have stacked 11 statistical measures [1] in a vector for each ECG and applied kernel-based SVMs to perform signal classification. The measures we have considered are:

- the standard deviation (SDNN),
- the standard deviation of the average RR interval calculated over 5 minutes (SDANN),
- the root mean square of successive differences (RMSSD),

- the mean of the 5-min standard deviation of the RR interval calculated over 24 hours (SDNNi),
- the standard deviation of successive differences (SDSD),
- the number of interval differences of successive RR intervals greater than 50 ms over the number of RR intervals (pNN50),
- the mean prediction error of the signal using local linear prediction (LLP),
- the total number of all RR intervals divided by the height of the histogram of all RR intervals measured on a discrete scale with bins of 78125 ms (TI),
- the baseline width of the minimum square difference triangular interpolation of the highest peak of the histogram of all NN intervals (TINN),
- the entropy of the signal,
- the autocorrelation of the signal.

### III. RESULTS

We have established several experimental configurations. We have applied SVM classification to a subset of the Fantasia database [26]. This database consists of twenty young (21 - 34 years old) and twenty elderly (68 - 85 years old) rigorously-screened healthy subjects which underwent 120 minutes of continuous supine resting while continuous ECG signals were collected. All subjects remained in a resting state in sinus rhythm while watching the movie Fantasia (Disney, 1940) to help maintain wakefulness [26]. In this study, we present results for the 10 heart beat time series consisting of five young and five elderly subjects.

In that framework, we also compared the SVM approach with the LVQ neural network [22]. LVQ is an autoassociative nearest neighbor classifier which classifies arbitrarily patterns into classes using an error correction encoding procedure related to competitive learning. The main idea is to cover the input space of samples with codebook vectors, each representing a region labeled with a class. A codebook vector can be seen as a prototype of a class member, localized in the center of a class or decision region (Voronoi cell) in the input space. A class can be represented by an arbitrarily number of codebook vectors, but one codebook vector represents one class only.

At first, leave-one-out cross-validation was used to evaluate the classifiers. More precisely, we have trained the classifiers with nine signals leaving one as a test signal. This was performed for all of the ten signals. In all cases the SVM-based categorization was 100% accurate while the LVQ classified correctly 83% of the signals.

A second experiment consists in investigating the robustness to noise of the SVM classifier. The original signals were corrupted by zero mean white Gaussian noise. The standard deviation of the noise was selected appropriately in order to obtain signal to noise ratios (SNR) between 5 and 0 dB. For each signal, 50 new signals were generated. Attribute vectors were created from the degraded signals and the classifiers were trained by a total number of 500 signals for each SNR level. 200 new test signals were

then created by the same procedure for each SNR level and the classifiers were evaluated on them. We have to mention that when important amount of noise is added to the signal, the statistical attributes cannot classify the signals by themselves. For instance, fig. 1 illustrates the modification in some statistical measures introduced by the noise. As it can be noticed, a value of 16 for the triangular index and a value of 0.03 for the SDANN separates the young from the elderly subjects (fig. 1(a)-(b)). On the other hand, when the original RR signals are corrupted by white Gaussian noise categorization by simple thresholding becomes impossible.

Table I summarizes the performances of the compared classifiers. As it can be seen, the SVM approach performs better than the LVQ neural network.

TABLE I

*Percentage of correctly classified signals using SVM and LVQ classifiers. In absence of noise, the LVQ classifies correctly 83% of the signals and the SVM classifier achieves 100% correct classification.*

Correct classification (%)		
SNR	LVQ	SVM
5 dB	70%	100%
4 dB	69%	99%
3 dB	65%	95%
2 dB	62%	89%
1 dB	60%	88%
0 dB	62%	70%

### IV. CONCLUSION

The categorization of the ECGs into two distinct groups according to their heart rate variability can be very accurate using an SVM in cases where standard methods fail to present a satisfactory categorization. Experiments comparing SVM classification of heart rate signals with the classifications obtained by other non-linear classifiers have also confirmed the effectiveness of the former methodology even in the presence of important amount of noise. A perspective of this study is the application of the SVM classifier to other ECG databases.

### REFERENCES

- [1] 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," *European Heart Journal*, vol. 17, pp. 354-381, 1996.
- [2] M. Kamath and E. Fallen, "Power spectral analysis of HRV: a noninvasive signature of cardiac autonomic functions," *Critical Reviews in Biomedical Engineering*, vol. 21, pp. 245-311, 1993.
- [3] R. Silipo, G. Deco, R. Vergassola, and C. Gremigni, "A characterization of HRV's nonlinear hidden dynamics by means of Markov models," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 8, pp. 978-986, 1999.
- [4] M. Ferrario, M. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 119-125, 2006.
- [5] D. Lake, "Renyi entropy measures for heart rate Gaussianity," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 21-27, 2006.

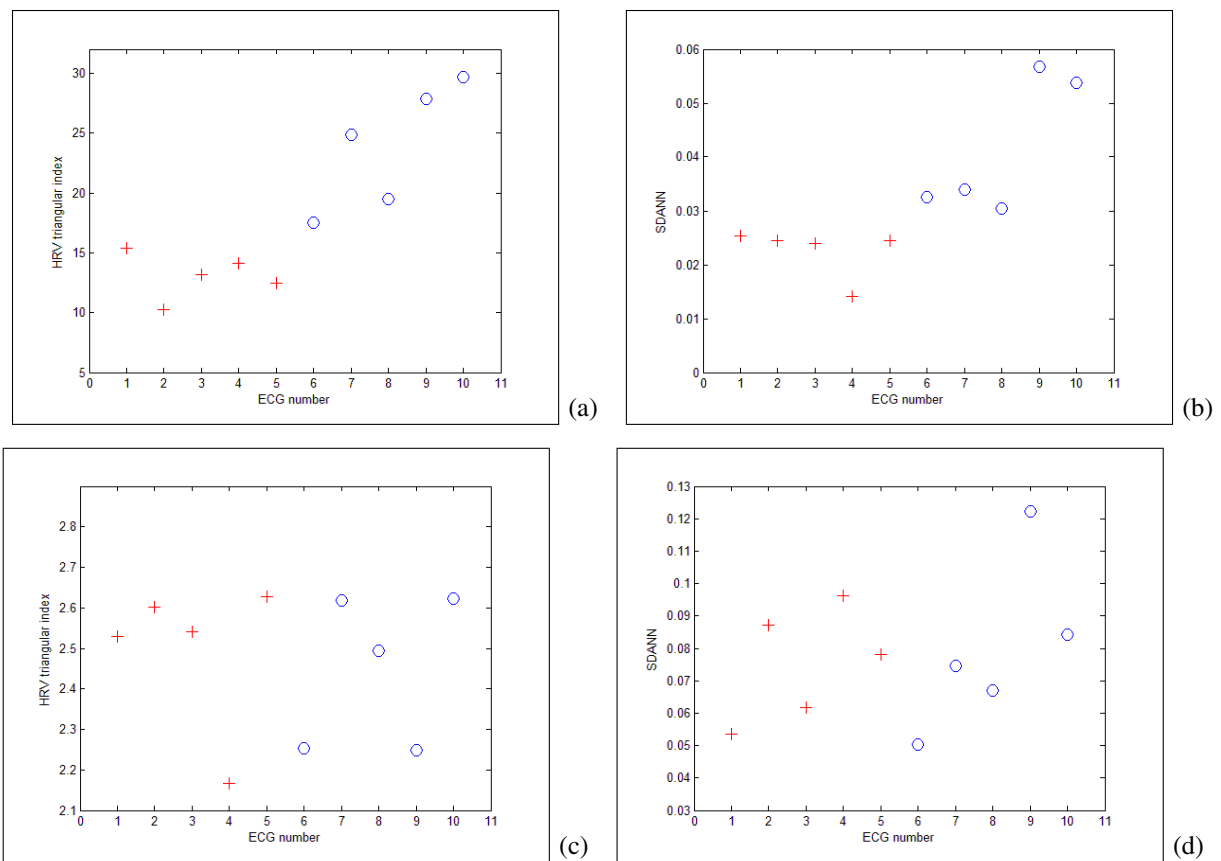


Fig. 1. (a) the triangular and (b) the SDANN statistical measures for 10 signals of the Fantasia database. (c)-(d) the same measures when the same RR signal is corrupted by zero mean white Gaussian noise resulting to a signal to noise ratio of 0 dB. Crosses represent young and circles represent elderly subjects. Notice that the measures do not categorize the signals if the noise degrades them significantly.

[6] D. Hoyer, B. Pompe, K. Chon, H. Hardhalt, C. Wicher, and U. Zwiener, "Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 4, pp. 584–592, 2005.

[7] K. Kiyono, Z. Struzik, N. Aoyagi, and Y. Yamamoto, "Multiscale probability density function analysis: Non-Gaussian and scale-invariant fluctuations of healthy human," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 95–102, 2006.

[8] R. Barbieri and E. Brown, "Analysis of heartbeat dynamics by point process adaptive filtering," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 4–12, 2006.

[9] B. Aysin, L. Chaparro, I. Gravé, and V. Shusterman, "Orthonormal basis partitioning and time frequency representation of cardiac rhythm dynamics," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 878–889, 2005.

[10] J. Mateo and P. Laguna, "Improved heart rate variability signal analysis from the beat occurrence times according to the IPFM model," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 8, pp. 997–1009, 2000.

[11] K. Solem, P. Laguna, and L. Sörnmo, "An efficient method for handling ectopic beats using the heart timing signal," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 13–20, 2006.

[12] C. Cortes and V. N. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 1–25, 1995.

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

[14] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[15] H. Byun and S. W. Lee, "A survey of pattern recognition applications of support vector machines," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 3, pp. 459–486, 2003.

[16] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[17] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542–1550, 2002.

[18] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 637–646, 1998.

[19] S. Li, T. Y. Kwok, I. Wai-Hang, and Y. Wang, "Fusing images with different focuses using support vector machines," *IEEE Transactions on Neural Networks*, vol. 15, no. 6, pp. 1555–1561, 2004.

[20] G. Pajares and J. M. de la Cruz, "On combining support vector machines and simulated annealing in stereovision matching," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 34, no. 4, pp. 1646–1657, 2004.

[21] A. Kampouraki, C. Nikou, and G. Manis, "Classification of heart rate signals using support vector machines," in *BioSignal*, Brno, Czech Republic, June 2006.

[22] T. K. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[23] B. Schölkopf, C. Burges, and A. Smola, *Advances in Kernel Methods: Support Vector Learning*. New York: MIT Press, 1999.

[24] R. Kondor and T. Jebara, "A kernel between sets of vectors," in *20<sup>th</sup> International Conference on Machine Learning (ICML)*, Washington DC, USA, 2003.

[25] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–202, 2001.

[26] Physiobank: Physiologic signal archives for biomedical research.