Modeling electrocardiogram using Yule-Walker equations and kernel machines

Maya Kallas*†, Clovis Francis*, Paul Honeine†, Hassan Amoud‡ and Cédric Richard§

*Laboratoire d'Analyse et de Surveillance des Systèmes (LASYS), Faculty of Engineering, Lebanese University, Lebanon
†Institut Charles Delaunay (UMR CNRS 6279), Université de Technologie de Troyes, France
‡Azm Center for Research in Biotechnology and its Applications, Lebanese University, Lebanon
§Laboratoire Fizeau (UMR CNRS 6525), Université de Nice Sophia-Antipolis, France

Abstract—One may monitor the heart normal activity by analyzing the electrocardiogram. We propose in this paper to combine the principle of kernel machines, that maps data into a high dimensional feature space, with the autoregressive (AR) technique defined using the Yule-Walker equations, which predicts future samples using a combination of some previous samples. A pre-image technique is applied in order to get back to the original space in order to interpret the predicted sample. The relevance of the proposed method is illustrated on real electrocardiogram from the MIT benchmark.

Index Terms—kernel machines, ECG signals, autoregressive model, nonlinear models, pre-image problem

I. INTRODUCTION

Heart disease remains the main cause of death nowadays. There are many different forms of heart disease, such as the heart failure, the narrowing or blockage of the coronary arteries. One way to monitor the heart is the electrocardiography. The record, which is called the electrocardiogram (ECG), shows the series of waves that relate to the electrical impulses which occur during each beat of the heart, and can be detected at the surface of the body. Many studies have been made in order to detect and classify the heart diseases of the ECG signals [1], [2], and [3].

An ECG is a time series corresponding to the electrical activity of the heart. The autoregressive (AR) model is one of the most useful technique for time series analysis [4]. Its concept is the prediction of a sample using a combination of some previous ones. Being a linear combination, the model is thus defined by the coefficients, or the weights in the linear expansion. In order to use such model, one has to identify its model parameters that are the coefficients and its order which is the number of the previous samples. A technique used to estimate its coefficients is the Yule-Walker equations. It is based on the evaluation of the autocorrelation function. Still, this linear technique is inappropriate for nonlinear systems such as modeling ECG.

In order to extend the AR to the case of nonlinear systems, we propose to apply the concept of the kernel methods. The main idea behind using the kernels is to map the data from an input space to a feature space using a nonlinear mapping function. In practice, it is not necessary to explicitly determine the mapping function, one can transform a linear algorithm into a nonlinear one by applying the *kernel trick* [5]. Initially introduced with Vapnik's Support Vector Machines (SVM) [6], the use of kernels proliferated in machine learning due to their low computational cost with high performances. Recently, some kernel-based methods were considered for the analysis and prediction of time series data [7], including the SVM for regression and kernel Kalman filter [8].

In this paper, we propose to combine the concept of the kernel machine with the simplicity of the AR model for modeling real ECG signal. The kernel AR is determined using the Yule-Walker equations in the feature space. Thus, parameters are estimated using the (lagged) expected kernels. To this end, we use a Gaussian kernel with the AR model estimated using the Yule-Walker equations. Eventhough, the kernel AR estimates the samples in some feature space. Therefore, being in a high dimensional space, one cannot interpret the sample. Thus, it is required to map back the data from the feature space to the input space where the sample can be interpreted. This necessitate the use of a pre-image technique to predict the future sample into a space where it can be observed. Many techniques have been presented in literature, in order to solve the pre-image problem. See [9] for a recent review. The paper's authors have solved the pre-image problem for biomedical signal processing in [10], [11], [12].

The rest of the paper is organized as follows: in the next section, we define the Yule-Walker equations used to estimate the coefficients of the AR model. In section III, we present the principle of the kernel methods in machine learning, and we extend the use of the AR model to the nonlinear systems. A pre-image technique is described in section IV in order to estimate the future sample. Section V illustrates the efficiency of the proposed technique on some real ECG signal taken from the MIT-BIH Normal Sinus Rhythm Database.

II. YULE-WALKER EQUATIONS FOR AR MODEL

A real-valued stochastic process x_i is said to be an *autoregressive process of order* p, denoted by AR(p) if there exists $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$, with $\alpha_p \neq 0$, and a white noise ϵ_i such that

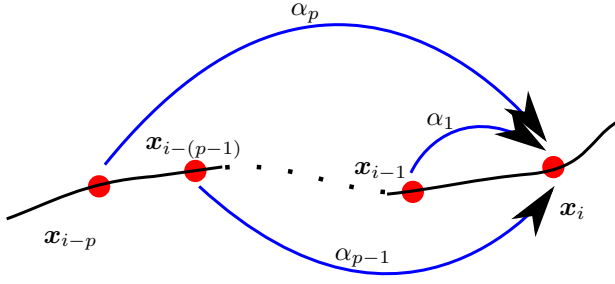$$x_i = \sum_{j=1}^{p} \alpha_j x_{i-j} + \epsilon_i. \qquad (1)$$

Fig. 1. Illustration of the AR concept, where the $\boldsymbol{x}_i$ is defined using the parameters α_k and the p previous $\boldsymbol{x}_{i-k}$.

Therefore, the set $x_{p+1}, x_{p+2}, \ldots, x_n$ defines a time series, described by the AR model of order p. The concept of the AR model is illustrated in figure 1. Several techniques in the literature allow to determine the parameters α_k. One of them is the resolution of the Yule-Walker equations. They entail the recursive computation of the autocorrelation function ρ of the AR(p)-process. The recursion $\rho(s) = \sum_{j=1}^{p} \alpha_j \rho(s - j)$, for lag $s > 0$ refers to the Yule-Walker equations. Having $\rho(-s) = \rho(s)$, and taking the empirical counterparts $r(j)$ of the autocorrelation $\rho(j)$, we get a more formal equation

$$\boldsymbol{r} = \boldsymbol{R}\boldsymbol{\alpha},$$

where $\boldsymbol{r} = [r(1) \ldots r(p)]^\top$, $\boldsymbol{\alpha} = [\alpha_1 \ldots \alpha_p]^\top$, and

$$\boldsymbol{R} = \begin{bmatrix} 1 & r(1) & \ldots & r(p-1) \\ r(1) & 1 & \ldots & r(p-2) \\ \vdots & & \ddots & \vdots \\ r(p-1) & r(p-2) & \ldots & 1 \end{bmatrix}.$$

When the $p \times p$ matrix $\boldsymbol{R}$ is invertible, we can rewrite the formal equation as $\boldsymbol{R}^{-1}\boldsymbol{r} = \boldsymbol{\alpha}$.

III. AR MODEL WITH KERNEL MACHINES

A. Kernel methods

The aforementioned techniques are used in the case of linear data. In order to extend their use to the nonlinear case, we propose the kernel machines. Kernel functions are typically viewed as providing an implicit mapping of points into a high-dimensional space, with the ability to gain much of the power of that space.

Let $\mathcal{X}$ be an input space with the canonical (Euclidean) dot product $x_i \cdot x_j$ for any $x_i, x_j \in \mathcal{X}$. Let $\kappa: \mathcal{X} \times \mathcal{X} \mapsto \mathcal{H}$ be a symmetric and continuous function, i.e., a kernel. A kernel is positive semi-definite if and only if a matrix K with entries $\kappa(x_i, x_j)$ for any finite subset of $\mathcal{X}$ is positive semi-definite, that is $\sum_{i,j} \alpha_i \alpha_j \kappa(x_i, x_j) \geq 0$ for all $\alpha_i, \alpha_j \in \mathbb{R}$ and all $x_i, x_j \in \mathcal{X}$. Based on the Moore-Aronszajn theorem [13], any positive semi-definite kernel guarantees the existence of a unique[1] feature space (or reproducing kernel Hilbert space) $\mathcal{H}$, where κ defines an inner product. Let $\Phi(\cdot)$ denotes the

[1]Unique, up to an isometry.

mapping function from the input space $\mathcal{X}$, into the feature space $\mathcal{H}$, then

$$\kappa(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}},$$

for any $x_i, x_j \in \mathcal{X}$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ denotes the corresponding inner product in $\mathcal{H}$, and $\| \cdot \|_{\mathcal{H}}$ its norm.

B. Kernel AR model

Using the concept of the kernel machines, and applying it to the data of some time series, each x_i of the series is mapped into the feature space using the mapping function $\Phi(\cdot)$, yielding to $\Phi(x_i)$. So in the case of the AR model, when extending it to the nonlinear systems, and applying the kernel machines, it will be defined in the feature space by

$$\Phi(x_i) = \sum_{j=1}^{p} \alpha_j \Phi(x_{i-j}) + \epsilon_i^\Phi.$$

Figure 2 illustrates the concept of the kernel AR model. Let μ be the mean of the mapped time series, namely

$$\mu = \mathbb{E}[\Phi(x_i)],$$

where $\mathbb{E}[\cdot]$ is the expectation of $\Phi(x_i)$ over all x_i's. Then,

$$\Phi(x_i) - \mu = \sum_{j=1}^{p} \alpha_j \Phi(x_{i-j}) + \epsilon_i^\Phi - \mu$$

$$= \sum_{j=1}^{p} \alpha_j \big(\Phi(x_{i-j}) - \mu \big) + \epsilon_i^\Phi - \Big(1 - \sum_{j=1}^{p} \alpha_j \Big) \mu.$$

With the expectation of both sides, the times series being assumed stationary, we get $(1 - \sum_{j=1}^{p} \alpha_j)\mu = \mathbb{E}[\epsilon_i^\Phi]$. By taking the centered data $(\Phi(x_{i-\tau}) - \mu)$, with lag $\tau > 0$, we get

$$\langle \Phi(x_i) - \mu, \Phi(x_{i-\tau}) - \mu \rangle_{\mathcal{H}} = \langle \epsilon_i^\Phi - \mathbb{E}[\epsilon_i^\Phi], \Phi(x_{i-\tau}) - \mu \rangle_{\mathcal{H}}$$

$$+ \sum_{j=1}^{p} \alpha_j \langle \Phi(x_{i-j}) - \mu, \Phi(x_{i-\tau}) - \mu \rangle_{\mathcal{H}}.$$

$$(2)$$

Using the kernel trick, we replace the dot product in the feature space by some kernel function, we define the *centered version* of a kernel $\kappa(\cdot, \cdot)$ with

$$\kappa_c(x_i, x_j) = \langle \Phi(x_i) - \mu, \Phi(x_j) - \mu \rangle_{\mathcal{H}}.$$

By analogy with the linear AR case, we assume that the noise ϵ_i^Φ and $\Phi(x_{i-\tau})$ are uncorrelated for every positive lag τ. Therefore, by taking the expectations of expression (2) and assuming the stationarity, we get for any $\tau \geq 1$:

$$\mathbb{E}[\kappa_c(x_i, x_{i-\tau})] = \sum_{j=1}^{p} \alpha_j \mathbb{E}[\kappa_c(x_{i-j}, x_{i-\tau})], \qquad (3)$$

where the notion of expected kernels is equivalent to the one recently studied in [14]. By considering all the lag values, expression (3) is written in matrix form

$$\boldsymbol{r}_\kappa = \boldsymbol{R}_\kappa \boldsymbol{\alpha},$$
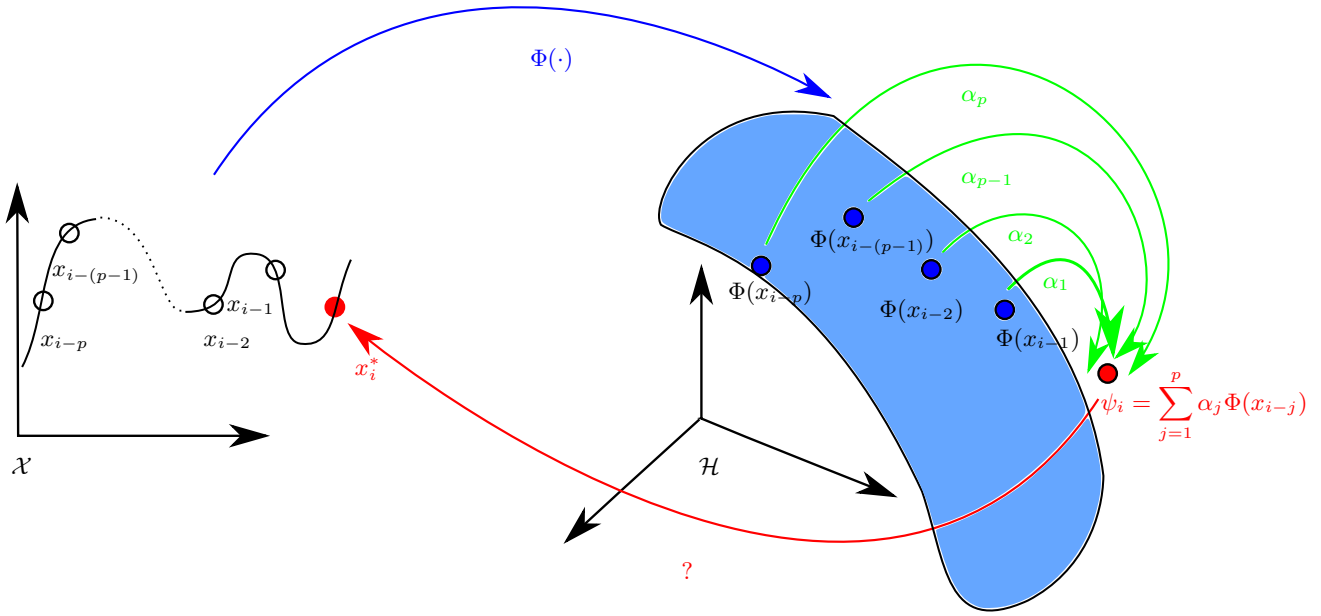
Fig. 2. Illustration of the Kernel AR model concept, for which, the data are mapped from $\mathcal{X}$ to $\mathcal{H}$, where the linear AR concept is applied on the mapped data. However, once ψ_i is evaluated, a mapping back from $\mathcal{H}$ to $\mathcal{X}$ is needed to interpret $\boldsymbol{x}_i$.

where

$$\boldsymbol{r}_\kappa = \begin{bmatrix} \mathbb{E}[\kappa_c(x_i, x_{i-1})] & \mathbb{E}[\kappa_c(x_i, x_{i-2})] & \cdots & \mathbb{E}[\kappa_c(x_i, x_{i-p})] \end{bmatrix}^\top,$$

and $\boldsymbol{R}_\kappa$ is the matrix described by the expected kernels with

$$\begin{bmatrix} \mathbb{E}[\kappa_c(x_i, x_i)] & \mathbb{E}[\kappa_c(x_{i-2}, x_{i-1})] & \cdots & \mathbb{E}[\kappa_c(x_{i-p}, x_{i-1})] \\ \mathbb{E}[\kappa_c(x_{i-1}, x_{i-2})] & \mathbb{E}[\kappa_c(x_i, x_i)] & \cdots & \mathbb{E}[\kappa_c(x_{i-p}, x_{i-2})] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[\kappa_c(x_{i-1}, x_{i-p})] & \mathbb{E}[\kappa_c(x_{i-2}, x_{i-p})] & \cdots & \mathbb{E}[\kappa_c(x_i, x_i)] \end{bmatrix}$$

The vector of coefficients $\boldsymbol{\alpha}$ is can be calculated by inverting the matrix $\boldsymbol{R}_\kappa$, with

$$\boldsymbol{\alpha} = \boldsymbol{R}_\kappa^{-1} \boldsymbol{r}_\kappa.$$

In practice, the expectations are estimated over a set of n available samples. The centered version of the kernel is determined using

$$\kappa_c(x_i, x_j) = \kappa(x_i, x_j) - \frac{1}{n} \sum_{k=1}^n \kappa(x_i, x_k) - \frac{1}{n} \sum_{k=1}^n \kappa(x_j, x_k)$$
$$+ \frac{1}{n^2} \sum_{k,k'=1}^n \kappa(x_{k'}, x_k).$$

IV. PRE-IMAGE TECHNIQUE AS A PREDICTION OF FUTURE SAMPLES

When the learning stage has been completed with some n available samples, the model parameters are estimated. The AR model in feature space can thus estimate the future sample by

$$\psi_i = \sum_{j=1}^p \alpha_j \, \Phi(x_{i-j}),$$

starting with $i = n + 1$. This evaluation of p samples in the feature space gives a sample in such space. Nonetheless, in a high space, this sample can not be interpreted. Thus in order to predict a future sample in the original space, one needs to map back ψ_i from the feature space into the input space. Nevertheless, a combination can lead to sample outside the image space, as shown in figure 2, thus, the sample may not have a pre-image, or if it exists it may not be unique. This is the pre-image problem, where one seeks to find an approximate solution in the input space, whose image by the nonlinear function is as close as possible to the sample estimated in the feature space. Several techniques have been proposed to solve this ill-posed problem. This includes a multidimensional scaling technique, a conformal map, and a learning scheme. See [9] for a formal definition of the pre-image problem, and a recent survey of the literature.

As aforementioned, the pre-image problem is solved when we can find an approximate solution x_i^* whose image $\Phi(x^*)$ is as close as possible to ψ_i. This is defined by the following optimization problem:

$$x_i^* = \arg\min_x \frac{1}{2} \left\| \Phi(x) - \sum_{j=1}^p \alpha_j \, \Phi(x_{i-j}) \right\|_{\mathcal{H}}^2.$$

In general, we can write it as

$$x_i^* = \arg\min_x J_i(x),$$

with

$$J_i(x) = -\sum_{j=1}^p \alpha_j \, \kappa(x_{i-j}, x) + \frac{1}{2} \kappa(x, x), \qquad (4)$$

where the term independent of x has been removed.

One technique used to solve such optimization problem is the fixed-point iterative technique. Based on the gradient

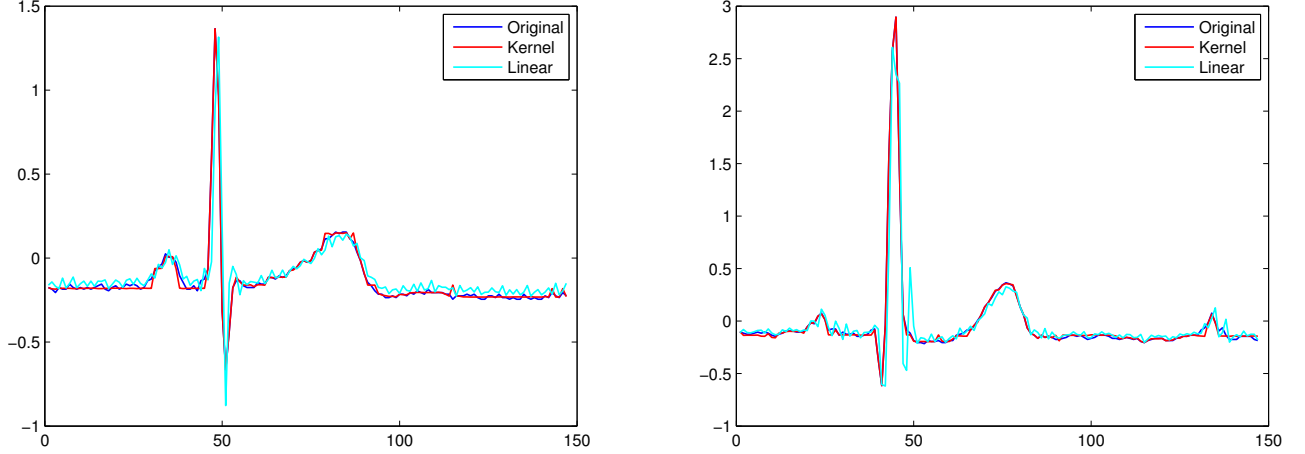| Signal number	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Mean of MSE over the 10 signals
Kernel AR	0.00056	0.00030	0.00021	0.00027	0.00014	0.00022	0.00018	0.00039	0.00021	0.00059	0.000313
Linear AR	0.0396	0.0091	0.0631	0.0210	0.0203	0.0268	0.0619	0.221	0.0169	0.0077	0.04874

Fig. 3. The modeling of 2 signals, using the kernel AR method, and the linear AR method. The blue signal presents the original one, the red is the one modeled using the kernel method, and the cyan is the one modeled using the linear AR.

descent scheme, and the type of kernel used, we can implement an iterative scheme to identify the pre-image [15]. The Gaussian kernel is the most used one. It is defined by

$$\exp(-\|x_i - x_j\|^2/2\sigma^2),$$

for any $x_i, x_j \in \mathcal{X}$, where σ defines its bandwidth parameter. The resulting cost function $J_i(x)$ is defined by

$$-\sum_{j=1}^{p} \alpha_j \exp(-\|x_{i-j} - x\|^2/2\sigma^2).$$

Taking the derivative of the above expression with respect to x, and setting it to zero, we get the fixed-point iterative expression

$$x_{i,t+1}^* = \frac{\sum_{j=1}^{p} \alpha_j \exp(-\|x_{i-j} - x_{i,t}^*\|^2/2\sigma^2)\, x_{i-j}}{\sum_{j=1}^{p} \alpha_j \exp(-\|x_{i-j} - x_{i,t}^*\|^2/2\sigma^2)}.$$

V. EXPERIMENTATIONS

We applied the proposed method to predict future samples of the ECG signals on some real ones taken from the MIT-BIH Normal Sinus Rhythm Database. The ECG signals are considered stationary, foe they are taken in a 1 minute of time. To this end, the gaussian kernel was used. The experiments were divided into two stages: learning and testing. 150 samples from each signal were used in the learning stage, where the order and the coefficients of the AR model and the bandwidth of the kernel were optimized. The same order for each signal

is taken for both the linear AR and the kernel AR models. The following 150 samples were used to test the relevance of the proposed kernel AR model, to the linear AR model where its coefficients are estimated using a linear Yule-Walker equations.

A comparison between the kernel AR method, and the linear AR model has been made. The comparison is evaluated based on the calculation of mean square error estimated on the 150 samples of the test stage. The mean square error (MSE) was estimated with

$$\varepsilon = \frac{1}{n} \sum_{i=n+1}^{2n} \|x_i^* - x_i\|^2,$$

where x_i^* is the predicted value at instant i, and x_i is the true value of the ECG signal at the same time. Table I shows the MSE for the kernel AR method and the linear AR method for the ECG signals, while the last column shows the mean of the MSE over the 10 signals. As we can see, the kernel AR model gives better results than the linear AR model. The MSE in the first case is approximately 100 times less than in the linear AR case.

Figure 3 shows two examples of modeling signals for normal cardiac rhythm, on the one hand, using the kernel AR technique estimated with the Yule-Walker equations in the feature space, and on the other, using the linear AR technique

evaluated with the Yule-Walker equations. The blue line is the original ECG signal, the red is the one modeled using the proposed kernel AR method, and the cyan is the one modeled using the linear AR. As we can see, the proposed method presents a more appropriate modeling than the linear AR method. The linear AR method fails to model the QRS complex of the heart beat, while the kernel AR method models the complex for each person based on his heart rate.

VI. CONCLUSION

Heart disease is one of the main death causes these days. Yet by analyzing and monitoring the ECG signals, one can predict when an arrythmia may occur. In this paper, we have presented a way to predict the evolution of the electrical activity of the heart using kernel machines. To this end, we have combined the concept of the AR model, which is a linear way to predict sample for time series, with the kernel methods in machine learning, where the data are mapped into a high dimensional feature space. We have defined the kernel AR in feature space using the Yule-Walker equations in order to estimate its coefficients. Once the future samples are estimated using this technique, a pre-image technique is required to map the data back to the input space, where the samples can be interpreted. The pre-image technique used was the fixed-point iterative method. The efficiency of the proposed method is illustrated on some real ECG signals taken from the MIT-BIH Normal Sinus Rhythm Database.

As for future work, we consider the estimation of the AR model using the partial autocorrelation function, or the Bayesian Information Criterion. Also, we intend to extend the use of the kernel machines to the autoregressive moving average or autoregressive integrated moving average model.

ACKNOWLEDGMENT

REFERENCES

[1] T. Linh, S. Osowski, and M. Stodolski, "On-line heart beat recognition using hermite polynomials and neuro-fuzzy network," *IEEE Transactions On Instrumentation And Measurement*, vol. 52, no. 4, August 2003.

[2] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions On Biomedical Engineering*, vol. 51, no. 4, Avril 2004.

[3] W. Kao, C. Yu, C. Shen, W. Chen, and P. Hsiao, "Electrocardiogram analysis with adaptive feature selection and support vector machines," in *IEEE Asia Pacific Conference on Circuits and Systems, APCCAS*, Singapore, 4-7 December 2006, pp. 1783–1786.

[4] M. Chevalier and Y. Grenier, "Autoregressive models with time-dependent log area ratios," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85.*, vol. 10, apr 1985, pp. 1049 – 1052.

[5] M. A. Aizerman, E. A. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," in *Automation and Remote Control*, no. 25, 1964, pp. 821–837.

[6] V. N. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, September 1998.

[7] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE trans. on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, March 2009.

[8] L. Ralaivola and F. D'alche-Buc, "Time series filtering, smoothing and learning using the kernel kalman filter," in *Proc. IEEE International Joint Conference on Neural Networks*, vol. 3, 2005, pp. 1449–1454.

[9] P. Honeine and C. Richard, "The pre-image problem in kernel-based machine learning," *IEEE Signal Processing Magazine, special issue on "dimensionality reduction via subspace and manifold learnin"*, vol. 28 (2), March 2011.

[10] M. Kallas, P. Honeine, C. Richard, H. Amoud, and C. Francis, "Non-linear feature extraction using kernel principal component analysis with non-negative pre-image," in *Proc. 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Buenos Aires, Argentina, 31 Aug. - 4 Sept. 2010.

[11] M. Kallas, P. Honeine, C. Richard, C. Francis, and H. Amoud, "Non-negative pre-image in machine learning for pattern recognition," in *19th European Signal Processing Conference*, Barcelona, Spain, 29 August - 2 September 2011.

[12] M. Kallas, P. Honeine, H. Amoud, and C. Francis, "Sur le problème de la pré-image en reconnaissance des formes avec contraintes de non-négativité," in *Colloque GRETSI'2011*, Bordeaux, France, 5-8 Septembre 2011.

[13] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[14] H. S. Anderson and M. R. Gupta, "Expected kernel for missing features in support vector machines," in *IEEE Workshop on Statistical Signal Processing*, Nice, France, 28-30 June 2011.

[15] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," in *Proceedings of the 1998 conference on Advances in neural information processing systems II*. Cambridge, MA, USA: MIT Press, 1999, pp. 536–542.