

Multi-Class ECG Beat Classification Based on a Gaussian Mixture Model of Karhunen-Loève Transform

Paolo Crippa, Alessandro Curzi, Laura Falaschetti, Claudio Turchetti
DII -- Dipartimento di Ingegneria dell'Informazione
Università Politecnica delle Marche
Via Brecce Bianche 12, I-60131 Ancona, Italy
e-mail: {p.crippa, a.curzi, l.falaschetti, c.turchetti}@univpm.it

Abstract — Cardiovascular diseases are one of the main causes of death around the world. Automatic classification of electrocardiogram (ECG) signals is of paramount importance in the unmanned detection of a wide range of heartbeat abnormalities. In this paper an effective multi-class beat classifier, based on a statistical identification of a minimum-complexity model, is presented. This methodology extracts from the ECG signal the multivariate relationships of its natural modes, by means of the separation property of the Karhunen-Loève transform (KLT). Then, it exploits an optimized expectation maximization (EM) algorithm to find the optimal parameters of a Gaussian mixture model, with the focus being in reducing the number of parameters. The resulting statistical model is thus based on the estimation of the multivariate probability density function (PDF) that characterizes each beat type. Based on the above statistical characterization a multi-class ECG classification was performed. The experiments, conducted on the ECG signals from the MIT-BIH arrhythmia database, demonstrated the validity and, considering the reduced model size, the excellent performance of this technique to classify the ECG signals into different disease categories.

Keywords - ECG; multi-class classification; Gaussian mixture model; KLT; expectation maximization; statistical model

I. INTRODUCTION

The electrocardiogram (ECG) signals [1], as well as electroencephalogram [2] or surface electromyogram signals [3], are some of the most important bio-electric signals that correspond to the electrical functionality of the human body. These signals could be considered as outputs of a nonstationary nonlinear system whose “input signals” are unknown. More in detail, the ECG signals reflect the performance and properties of the human heart, thus their analysis provides critical information on the cardiac function of patients.

The incidence of cardiovascular diseases (CVDs) have increased in recent years representing the biggest cause of deaths worldwide. More than 17 million people died from CVDs in 2008 and more than 3 million of these deaths occurred before the age of 60. The percentage of premature deaths from CVDs ranges from 4% in high-income countries to 42% in low-income countries, leading to growing inequalities in the occurrence and outcome of CVDs between countries and populations [4]. One of the complication of CVD among many others is atrial and ventricular arrhythmias which occur due to cardiac rhythm disturbances. Arrhythmia is a collective term for a heterogeneous group of conditions in which there would be abnormal electrical activity. Sometimes cardiac arrhythmia is life threatening and causes medical emergencies, sometimes it may not cause symptoms or it may give rise to palpitations. Arrhythmias like ventricular fibrillation and flutter are life threatening

medical emergencies which result in cardiac arrest, hemodynamic collapse and sudden cardiac death.

Cardiac disease conditions can be diagnosed by identifying abnormal heartbeats in the ECG signal. Therefore long-term recordings of the ECG signal are performed in healthcare applications such as clinical monitoring or telemedicine using the popular Holter recorders. Holter ambulatory systems record at least 24 hours of heart activity, resulting in data that contain thousands of heartbeats. The analysis is usually performed offline by cardiologists by analyzing ECG visually on paper or on a screen. Because of the high number of beats to evaluate and the complexity of their shape, this task is very time consuming and reliable visual inspection is difficult. This leads numerous scientists to study the automatic detection problem [5]–[14]. In particular, computer-aided multi-class classification of pathological beats is of paramount importance to perform correct diagnosis as reported in very recent works [11], [15]–[26] where several classification methods have been used such as support vector machine (SVM) [6], [19], [21]–[23], [27], neural networks [15], [16], [28], [29], hybrid neuro-fuzzy systems [30]–[32], as well as the powerful dynamical models such as Bayesian networks [33], [34] and hidden Markov models [35].

To this end, we propose an automatic classifier based on the identification of a statistical model of the ECG signal [36]. Indeed, the signals to be modeled are inherently stochastic, in that they contain unpredictable components not only due to the presence of noise. ECG signals, even if they

belong to the same individual, are well-known examples of stochastic signals.

The issue of selecting an optimal set of relevant features for the ECG signal plays an important role in beat classification. It is well-known that the performance of a classifier in general deteriorates when some of the selected features are correlated. Therefore the multi-class beat classification turns out to be a problem of classification with the smallest number of uncorrelated extracted features capable of separating the classes with a minimum error.

Thus, we opted for features that are automatically selected from the data properties themselves, and not fixed *a priori*. Among available methods that can be used to extract different features from the same raw data we chose the Karhunen-Loève transform (KLT), because it produces uncorrelated features whose number can be arbitrarily chosen according to the desired model accuracy, still maintaining optimal signal reconstruction capability [37].

Based on a stochastic setting, the technique we propose initially extracts from the signals the multivariate relationships existing between their natural modes. This was obtained by using the properties of spectral separability of the information contained in the KLT. Then it approximates the multivariate probability density function (PDF) of the KLT components with a Gaussian mixture model (GMM), and performs the model identification by finding the *maximum likelihood* of the mixture model parameters. To this end an efficient implementation of the expectation maximization (EM) algorithm, as presented in [38], is employed. Although other techniques can be employed for the above task, (e.g. the use of a self-organizing mixture network [39]–[41]), this algorithm appeared to be the most suitable for our application, as it also automatically selects the minimum number of Gaussians per mixture in order to jointly optimize the information loss due to the modeling error together with the model complexity. Having a limited number of model parameters is useful because it facilitates the possible implementation of the proposed technique in resource-constrained portable devices.

At the end of the identification process a statistical model of the different beat types is obtained, and a multi-class ECG classification can be performed, capable of distinguishing normal beats from several different types of arrhythmia.

Fig. 1 reports the flow chart of the overall multi-class ECG beat classification methodology presented in this paper.

For the classification six types of ECG beats were considered: normal, Right Bundle Branch Block (RBBB), Left Bundle Branch Block (LBBB), Premature Ventricular Contraction (PVC), Atrial Premature Contraction (APC) and paced beats.

The algorithm was validated through the baseline MIT-BIH arrhythmia database [42]. 89807 beats were extracted from which 44905 (50%) were used for training and the remaining 44902 beats for testing purposes. The proposed technique achieved an overall accuracy of more than 95% ad its performance was compared with state-of-the-art SVM multi-class classifier [19], [21], [22].

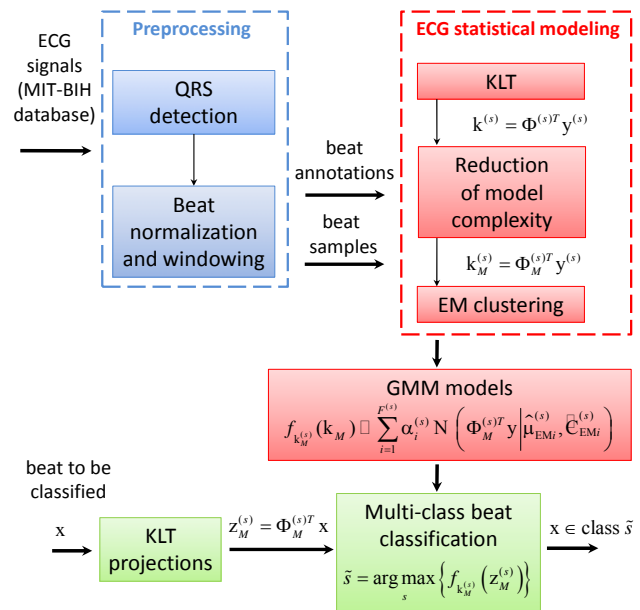


Figure 1. Flowchart of the proposed ECG beat classifier.

Due the large amount of data necessary for building an accurate statistical model for each of the considered beat classes, we opted for a “class-oriented” evaluation scheme, as adopted in most works, such as [43]–[45]. The heartbeats segments were extracted from records of the MIT-BIH arrhythmia database and were clustered, based on the categorization of the heartbeats. After randomization, a certain fraction of each cluster (i.e., class) was selected as the training dataset and the remaining heartbeats were used as the testing dataset.

Although the “class-oriented” evaluation could lead to optimistic results, because the inter-beat variations in ECG characteristics are reduced in such tests as the training and testing datasets contain heartbeats from the same subjects, i.e., a similar beat might be included in both sets, the “subject oriented” choice, used in other recent works [9], [11], [23], [25], [46], [47], could not be adopted here due the reduced number of available beats which is insufficient for building a reliable overall statistical model.

This paper is organized as follows. Section II presents the statistical model and Section III discusses the multi-class beat classification. Section IV describes the data set and its preprocessing. Section V reports the experimental results of the classifier on the database signals and comments the performance. Finally, conclusions are drawn in Section VI.

II. ECG SIGNAL STATISTICAL MODEL

Let us consider the ECG signal of finite length $y[n]$, $n = 0, \dots, N-1$, as an observation of the $N \times 1$ real random vector $y = [y_1, \dots, y_N]^T$ whose covariance function is given by

$$C_{yy} = E \left\{ (y - \mu_y)(y - \mu_y)^T \right\} \quad (1)$$

where the symbol $E\{\cdot\}$ denotes the expectation and $\mu_y = E\{y\}$.

It is well known that an orthonormal set [48] $\{\phi_1, \dots, \phi_N\}$, can be derived as a solution of the eigenvector equations

$$C_{yy}\phi_j = \lambda_j\phi_j, \quad j = 1, \dots, N \quad (2)$$

which can be put in matrix form as

$$C_{yy} = \Phi\Lambda\Phi^T \quad (3)$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$, $\Phi = [\phi_1, \dots, \phi_N] \in \mathbb{R}^{N \times N}$, and $\phi_j \in \mathbb{R}^{N \times 1}$ are such that

$$\phi_i^T \phi_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (4)$$

The Karhunen-Loève transform (KLT) of y is defined as

$$k = \Phi^T y \quad (5)$$

and the inverse transform is

$$y = \Phi k \quad (6)$$

where $k = [k_1, \dots, k_N]^T$ is the transformed random vector with

$$\mu_k = E\{k\} = \Phi^T \mu_y \quad (7)$$

The main property of the KLT is that in this expansion the coefficients, k_j , $j = 1, \dots, N$, form an orthogonal sequence of random variables, so that the covariance matrix of the random vector k is diagonal and given by

$$C_{kk} = E\{(k - \mu_k)(k - \mu_k)^T\} = \Lambda \quad (8)$$

as it can be easily verified by using (3) and (5). This effect, known as *decorrelation*, is a desirable property in defining a statistical model for a class of signals as it reduces the sparseness of random data to be modeled.

From the linear transform (6) it is easy to verify that the PDF $f_y(y)$ of y is related to the PDF $f_k(k)$ of k by the following relationship

$$f_y(y) = \frac{f_k(\Phi^T y)}{|\Phi|}, \quad (9)$$

where $|\Phi|$ is the magnitude of the determinant of the matrix Φ . The equalities $|\Phi| = 1$ and

$$f_y(y) = f_k(\Phi^T y) \quad (10)$$

hold, being Φ a unitary matrix.

This means that the problem of estimating the multivariate PDF of output data can be reduced to the density estimation of KLT components. To this end, given a set of observations $Y = \{y^{(\ell)}, \ell = 1, \dots, L\}$ to which corresponds the set $K = \{k^{(\ell)}, \ell = 1, \dots, L\}$ of KLT vectors, let us approximate the multivariate PDF $f_k(k)$ with an F -component Gaussian mixture

$$g(k|\Theta) = \sum_{i=1}^F \alpha_i N(k|\mu_i, C_i) \quad (11)$$

where α_i , $i = 1, \dots, F$ are the mixing weights, and

$$N(k|\mu_i, C_i) = \frac{(2\pi)^{-N/2}}{\sqrt{|C_i|}} \exp\left\{-\frac{1}{2}(k - \mu_i)^T C_i^{-1} (k - \mu_i)\right\} \quad (12)$$

represents the density of a Gaussian distribution with mean μ_i and covariance matrix C_i . It is worth noting that α_i , must satisfy $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^F \alpha_i = 1$. Θ is the set of parameters needed to specify the Gaussian mixture defined as

$$\Theta = \{\alpha_1, \dots, \alpha_F, \Theta_1, \dots, \Theta_F\}, \quad (13)$$

where

$$\Theta_i = \{\mu_i, C_i\} \quad (14)$$

It is well known that the *maximum likelihood* (ML) estimate of Θ ,

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} \left\{ \log [g(K|\Theta)] \right\} \quad (15)$$

is difficult to find because (15) contains the log of the sum. The usual choice for solving ML estimate of the mixture parameters is the EM algorithm.

This algorithm is based on a set $H = \{h^{(1)}, h^{(2)}, \dots, h^{(L)}\}$ of L labels associated with the L observations, each label being a binary vector $h^{(\ell)} = [h_1^{(\ell)}, h_2^{(\ell)}, \dots, h_F^{(\ell)}]$, where $h_i^{(\ell)} = 1$ and $h_l^{(\ell)} = 0$ for all $l \neq i$, means that the KLT vector $k^{(\ell)}$ was generated by the i -th Gaussian component $N(k|\mu_i, C_i)$.

The EM algorithm is based on the interpretation of K as incomplete data and H as the missing part of the complete data $X = \{K, H\}$. The complete data log-likelihood, i.e. the log-likelihood of X as though H was observed, is

$$\log [g(K, H|\Theta)] = \sum_{\ell=1}^L \sum_{i=1}^F h_i^{(\ell)} \log [\alpha_i N(k^{(\ell)}|\mu_i, C_i)] \quad (16)$$

In general the EM algorithm computes a sequence of parameter estimates $\{\hat{\Theta}(p), p = 0, 1, \dots\}$ by iteratively performing two steps:

- *Expectation step*: computes the expected value of the complete log-likelihood, given the training set K of KLT vectors and the current parameter estimate $\hat{\Theta}(p)$. The result is the so-called *auxiliary function*

$$Q(\Theta|\hat{\Theta}(p)) = E\left\{ \log [g(K, H|\Theta)] \mid K, \hat{\Theta}(p) \right\} \quad (17)$$

- *Maximization step*: update the parameter estimate

$$\hat{\Theta}(p+1) = \arg \max_{\Theta} \left\{ Q(\Theta|\hat{\Theta}(p)) \right\} \quad (18)$$

by maximizing the Q -function.

However, the EM algorithm for mixture estimation has several drawbacks: *i*) it is sensitive to initialization so it

requires careful initialization, and *ii*) it is not capable of selecting the number of components F .

Recently, *Figuereido et al.* [38] suggested an unsupervised algorithm for learning a finite mixture model from multivariate data, that overcomes the limitations of the standard EM approach.

This algorithm integrates both model estimation and component selection, i.e. the ability of choosing the best number of mixture components F according to a predefined minimization criterion, in a single framework. It is able to perform an automatic component annihilation directly within the maximization step of the EM iterations. Starting from the conventional equation for the maximization step defined in (18), the algorithm is modified as follows.

Compute

$$\hat{\alpha}_i(p+1) = \frac{\max\left\{0, \left(\sum_{i=1}^L u_i^{(p)}\right) - N\right\}}{\sum_{i=1}^F \max\left\{0, \left(\sum_{i=1}^L u_i^{(p)}\right) - N\right\}} \quad \text{for } i = 1, \dots, F \quad (19)$$

where the coefficients $u_i^{(p)}$, defined as

$$u_i^{(p)} = P\left(h_i^{(p)} = 1 | y^{(p)}\right), \quad (20)$$

represent the conditional expectations of the H elements and are obtained during the expectation step of EM algorithm.

For every i such that $\hat{\alpha}_i(p+1) > 0$, estimate $\hat{\Theta}_i$

$$\hat{\Theta}_i(p+1) = \arg \max_{\Theta_i} \left\{ Q\left(\Theta | \hat{\Theta}(p)\right) \right\}. \quad (21)$$

Note that (19) should be thought of as discriminative rules for reducing the amount of mixture components to be selected for the final model. It is worth noting that the parameters Θ in (21) are no longer computed simultaneously, instead they are updated sequentially by means of a component-wise version of the EM. Furthermore, it can be verified that starting from a large value of F , compared to the optimal number of mixture components, the overall model training procedure shows a sensible gain in robustness against bad initialization.

Thus, once the F -component Gaussian mixture is estimated, we can write

$$f_y(y) = f_k(\Phi^T y) \square g(\Phi^T y | \hat{\Theta}_{EM}), \quad (22)$$

where $\hat{\Theta}_{EM}$ is the final estimation of $\hat{\Theta}$ and this represents the ECG signal statistical model.

A. Reduction of Model Complexity

A further property of the KLT is that it is the most efficient representation of the random process if the expansion is truncated to $M < N$ orthonormal basis functions. Thus this property is useful in order to reduce the model complexity. To this end let us rewrite (6) as

$$y = \Phi k = \sum_{i=1}^N k_i \phi_i = \sum_{i=1}^M k_i \phi_i + \sum_{i=M+1}^N k_i \phi_i = x_M + \eta_y, \quad (23)$$

where

$$x_M = \sum_{i=1}^M k_i \phi_i = \Phi_M k_M, \quad (24)$$

with $\Phi_M = [\phi_1, \dots, \phi_M] \in \square^{N \times M}$, $k_M \in \square^M$, is the truncated expansion, and

$$\eta_y = \sum_{i=M+1}^N k_i \phi_i, \quad (25)$$

is the error or residual. It can be shown [48] that the minimum mean square error $E_M = E\{\eta_y^T \eta_y\}$, subject to the constraints $\phi_i^T \phi_i = 1$, $i = M+1, \dots, N$, is given by

$$E_M = E\left\{ (y - x_M)^T (y - x_M) \right\} = \sum_{i=M+1}^N \tilde{\lambda}_i, \quad (26)$$

where $\tilde{\lambda}_i$ represent the $N - M$ Lagrange multipliers of the constrained minimization, given by

$$\tilde{\lambda}_i = \lambda_i + \phi_i^T \mu_y \mu_y^T \phi_i, \quad i = M+1, \dots, N, \quad (27)$$

and λ_i is the eigenvalue corresponding to the eigenvector ϕ_i . In particular once the $\tilde{\lambda}_i$ are arranged in decreasing order, the error E_M decreases monotonically as the index M increases towards N .

In order to evaluate the impact of truncation on probability, let us start from the definition of mean square error

$$E_M = E\left\{ \|y - x_M\|^2 \right\} \square \int_{\square^N} \|y - x_M\|^2 p_y(y) dy, \quad (28)$$

where $\|\cdot\|$ is the norm of a vector, that is $\|v\| = (v^T v)^{1/2}$

For any $\varepsilon > 0$ it results

$$\begin{aligned} E\left\{ \|y - x_M\|^2 \right\} &\geq \int_{\|y - x_M\| > \varepsilon} \|y - x_M\|^2 p_y(y) dy \geq \\ &\geq \varepsilon^2 \int_{\|y - x_M\| > \varepsilon} p_y(y) dy = \varepsilon^2 P\left\{ \|y - x_M\| > \varepsilon \right\} \end{aligned} \quad (29)$$

where $P\{A\}$ denotes the probability of the set A . Hence one gets

$$P\left\{ \|y - x_M\| > \varepsilon \right\} \leq \frac{E\left\{ \|y - x_M\|^2 \right\}}{\varepsilon^2} = \frac{E_M}{\varepsilon^2}. \quad (30)$$

This result shows that the truncated expansion converges in probability [49] to the complete expansion as M tends to N meaning that the probability of the error monotonically decreases, as the truncated expansion tends to the complete expansion.

The truncation error E_k on the k -vector can be estimated as follows. Let us define the truncated vector k_T as

$$k_T = \begin{pmatrix} k_M \\ 0 \end{pmatrix} \in \square^N. \quad (31)$$

Thus we have

$$\begin{aligned} E_k &= E\left\{ \|k - k_T\|^2 \right\} = E\left\{ (k - k_T)^T (k - k_T) \right\} = \\ &= E\left\{ (k - k_T)^T \Phi^T \Phi (k - k_T) \right\} = E\left\{ \|\Phi(k - k_T)\|^2 \right\}, \end{aligned} \quad (32)$$

since $\Phi^T \Phi = I$. The term on the right-hand side becomes

$$E \left\{ \left\| \Phi \mathbf{k} - \Phi \mathbf{k}_T \right\|^2 \right\} = E \left\{ \left\| \mathbf{y} - [\Phi_M, \Phi_\eta] \begin{pmatrix} \mathbf{k}_M \\ \mathbf{0} \end{pmatrix} \right\|^2 \right\} = E \left\{ \left\| \mathbf{y} - \Phi_M \mathbf{k}_M \right\|^2 \right\} \quad (33)$$

where $\Phi_M = [\phi_1, \dots, \phi_M]$ and $\Phi_\eta = [\phi_{M+1}, \dots, \phi_N]$, so that it results $E_k = E_M$.

Furthermore the same approach used in deriving (30) yields

$$P \left\{ \left\| \mathbf{k} - \mathbf{k}_T \right\| > \varepsilon \right\} \leq \frac{E_k}{\varepsilon^2} \quad (34)$$

which combined with (26) gives an upper bound for the probability of the residual $\eta_k = \mathbf{k} - \mathbf{k}_T$

$$P \left\{ \left\| \mathbf{k} - \mathbf{k}_T \right\| > \varepsilon \right\} \leq \frac{\sum_{i=M+1}^N \tilde{\lambda}_i}{\varepsilon^2} \quad (35)$$

Although (34) merely shows that the truncation \mathbf{k}_T converges in probability to the complete vector \mathbf{k} , as it is well known that the convergence in probability implies convergence in distribution [49], we conclude that the pdf $f_{\mathbf{k}_T}(\mathbf{k}_T)$ converges to $f_{\mathbf{k}}(\mathbf{k})$.

In conclusion, with this result in mind, once the set $\{\phi_1, \dots, \phi_M\}$ of eigenvectors with the M largest $\tilde{\lambda}_i$ has been chosen, the statistical model used for classification is given by the projection

$$\mathbf{k}_M = \Phi_M^T \mathbf{y} \quad (36)$$

In such a way, the dimensionality M of the problem of estimating the multivariate PDF $f_{\mathbf{k}_M}(\mathbf{k}_M)$ is reduced, with negligible error in probability distribution.

III. MULTI-CLASS BEAT CLASSIFICATION

Now, to face multi-class beat classification, we consider S classes, corresponding to S different beat types. For each class s with $s = 1, \dots, S$, and $\mathbf{y}^{(s)}$ belonging to the s -class, we perform the KLT transforms $\mathbf{y}^{(s)} = \Phi^{(s)} \mathbf{k}^{(s)}$, $\mathbf{k}^{(s)} = \Phi^{(s)T} \mathbf{y}^{(s)}$, $\mathbf{k}_M^{(s)} = \Phi_M^{(s)T} \mathbf{y}^{(s)}$, by following the steps illustrated in the previous section. Then we obtain the statistical model for $\mathbf{k}_M^{(s)}$ as in (22),

$$f_{\mathbf{y}^{(s)}}(\mathbf{y}) \Delta \mathbf{y} = f_{\mathbf{k}^{(s)}}(\Phi^{(s)T} \mathbf{y}) \Delta \mathbf{k} \square f_{\mathbf{k}_M^{(s)}}(\mathbf{k}_M) \Delta \mathbf{k}_M \quad (37)$$

where

$$f_{\mathbf{k}_M^{(s)}}(\mathbf{k}_M) \square g \left(\Phi_M^{(s)T} \mathbf{y} \middle| \hat{\Theta}_{EM}^{(s)} \right) = \sum_{i=1}^{F^{(s)}} \alpha_i^{(s)} N \left(\Phi_M^{(s)T} \mathbf{y} \middle| \hat{\mu}_{EMi}^{(s)}, \hat{\Sigma}_{EMi}^{(s)} \right) \quad (38)$$

and $\Delta \mathbf{y}$, $\Delta \mathbf{k}$, $\Delta \mathbf{k}_M$ are volumes in the \mathbf{y} , \mathbf{k} , and \mathbf{k}_M spaces, respectively.

Now consider the generic beat \mathbf{x} to be classified, compute its projections over the KLT bases $\Phi^{(s)}$, $s = 1, \dots, S$

$$\mathbf{z}^{(s)} = \Phi^{(s)T} \mathbf{x} \quad , \quad s = 1, \dots, S \quad (39)$$

and define the score χ_s of the testing beat as

$$\chi_s = f_{\mathbf{k}_M^{(s)}}(\mathbf{z}_M^{(s)}) \quad (40)$$

where $\mathbf{z}_M^{(s)} \in \mathbb{R}^M$ are the first M components of $\mathbf{z}^{(s)}$:

$$\mathbf{z}_M^{(s)} = \Phi_M^{(s)T} \mathbf{x} = [z_1^{(s)}, \dots, z_M^{(s)}]^T \quad (41)$$

The class \tilde{s} to which \mathbf{x} belongs is the one that gives the maximum score χ_s

$$\tilde{s} = \arg \max_s \{ \chi_s \} \quad (42)$$

IV. DATASET

In order to test the effectiveness of the proposed classifier in the categorization of different beat types, suitable datasets were prepared to both train the model and test the classification accuracy.

Starting from the MIT-BIH arrhythmia database [42], which is widely regarded as a benchmark for arrhythmia detection and classification tasks, we selected a number of records and used half of the beats for training, retaining the other half for tests.

A. Materials

The original database contains 48 half-hour two-lead ambulatory ECG signals, recorded from 47 subjects in the late 1970's. The signals have then been filtered in the 0.1-100 Hz band and digitized at 360 Hz with 11 bit resolution with a 5 μ V least-significant-bit value. Each beat was carefully annotated and checked, and the QRS complex position also recorded together with the beat category annotation. Two lead signals were recorded but only one (usually lead A, a modified limb lead II, or MLII) used in our experiments, which also made use of the category annotations as the ground truth for performance evaluation.

Of all the available annotations, we made use of only the six main categories, for which a sufficiently high number of observations were present, i.e., (a) Normal, (b) Right Bundle Branch Block (RBBB), (c) Left Bundle Branch Block (LBBB), (d) Premature Ventricular Contraction (PVC), (e) Atrial Premature Contraction (APC) and (f) Paced beats.

In this work we chosen a "class-oriented" evaluation scheme, such as in [43]–[45], because this strategy allows to obtain a reliable and accurate statistical model for each type of arrhythmia (class) also in presence of a reduced number of available beats, i.e. when facing to small ECG databases.

The number of beats used for each class is reported in Table I: as it can be easily seen, this number varies ranging from 2384 for APC to 63625 for Normal beats. For each class, a random selection of 50% of the available beats was employed for training the models, while the remaining 50% had been set apart for evaluation during the testing phase. As a result a total of 89807 beats have been used, of which 44905 (44902) for the training (testing).

TABLE I. CONSISTENCY OF THE ECG BEAT DATABASE USED FOR EXPERIMENTAL EVALUATION

Class	Train	Test
<i>Normal</i>	31813	31812
<i>RBBB</i>	3540	3539
<i>LBBB</i>	2298	2298
<i>PVC</i>	2564	2564
<i>APC</i>	1192	1192
<i>Paced</i>	3498	3497
Total	44905	44902

B. ECG Signal Segmentation

In order to perform the KLT, fixed-length signal vectors are needed. The ECG signal was thus segmented in 200 ms windows (72 samples) centered around the R wave point as recorded in the annotated files and as suggested in [50]. Of course, a real application would need a QRS detector to automatically find this point, but many well-established algorithms exist to this end.

V. EXPERIMENTAL RESULTS

In this section we present the experimental results obtained with the proposed multi-class ECG signal classifier.

The signals used for the experiments were extracted from the MIT-BIH arrhythmia database as described earlier. The set of the KLT input signals was thus populated by L vectors, with L given in Table I according to the class being trained, of $N = 72$ samples of windowed signal data, with no further preprocessing applied. For every beat belonging to the testing set of each class, we then run our classifier and counted the number of occurrences of each recognized type, so as to obtain a confusion matrix.

Several experiments were performed by varying the number of KLT components retained in the GMM model. A selection of the resulting six-classes confusion matrices are reported in Tables II, III, and IV where 5, 10, and 20 KLT components, respectively, have been considered.

In order to obtain more compact figures related to the performance of the algorithm, the standard set of performance indices was also extracted from the confusion matrices, namely we computed the sensitivity, specificity, precision and accuracy, defined as

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad , \quad (43)$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad , \quad (44)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad , \quad (45)$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad , \quad (46)$$

TABLE II. CONFUSION MATRIX FOR THE PROPOSED ECG BEAT CLASSIFIER OBTAINED BY CONSIDERING $M = 5$ KLT COMPONENTS

Input	Recognized					
	<i>Normal</i>	<i>RBBB</i>	<i>LBBB</i>	<i>PVC</i>	<i>APC</i>	<i>Paced</i>
<i>Normal</i>	28416	239	96	412	2647	2
<i>RBBB</i>	25	3311	0	33	170	0
<i>LBBB</i>	31	0	2243	17	5	2
<i>PVC</i>	86	10	2	2437	12	17
<i>APC</i>	141	67	0	12	972	0
<i>Paced</i>	3	1	7	9	9	3468

TABLE III. CONFUSION MATRIX FOR THE PROPOSED ECG BEAT CLASSIFIER OBTAINED BY CONSIDERING $M = 10$ KLT COMPONENTS

Input	Recognized					
	<i>Normal</i>	<i>RBBB</i>	<i>LBBB</i>	<i>PVC</i>	<i>APC</i>	<i>Paced</i>
<i>Normal</i>	29710	51	56	302	1687	6
<i>RBBB</i>	16	3329	0	15	179	0
<i>LBBB</i>	5	0	2272	15	5	1
<i>PVC</i>	88	1	16	2455	4	0
<i>APC</i>	110	69	0	11	1002	0
<i>Paced</i>	3	0	4	13	0	3477

TABLE IV. CONFUSION MATRIX FOR THE PROPOSED ECG BEAT CLASSIFIER OBTAINED BY CONSIDERING $M = 20$ KLT COMPONENTS

Input	Recognized					
	<i>Normal</i>	<i>RBBB</i>	<i>LBBB</i>	<i>PVC</i>	<i>APC</i>	<i>Paced</i>
<i>Normal</i>	30521	0	1	260	1028	2
<i>RBBB</i>	47	3274	0	35	183	0
<i>LBBB</i>	17	0	2231	50	0	0
<i>PVC</i>	83	0	0	2477	4	0
<i>APC</i>	250	85	0	4	853	0
<i>Paced</i>	5	0	0	22	0	3470

where TP are the true positives (the diagonal elements of the confusion matrix), FN the false negatives (the sum of the other elements on the same row of the confusion matrix), FP the false positives (the sum of the other elements on the same column of the confusion matrix), and TN the true negatives (the sum of the elements on the other rows and columns of the confusion matrix). Results are reported in Fig. 2 (a)–(d), where different numbers of eigenvector projections M were retained in the GMM model, and also reported in Table V for the case of $M = 10$ eigenvectors, which we deemed to be the best compromise between all the indices.

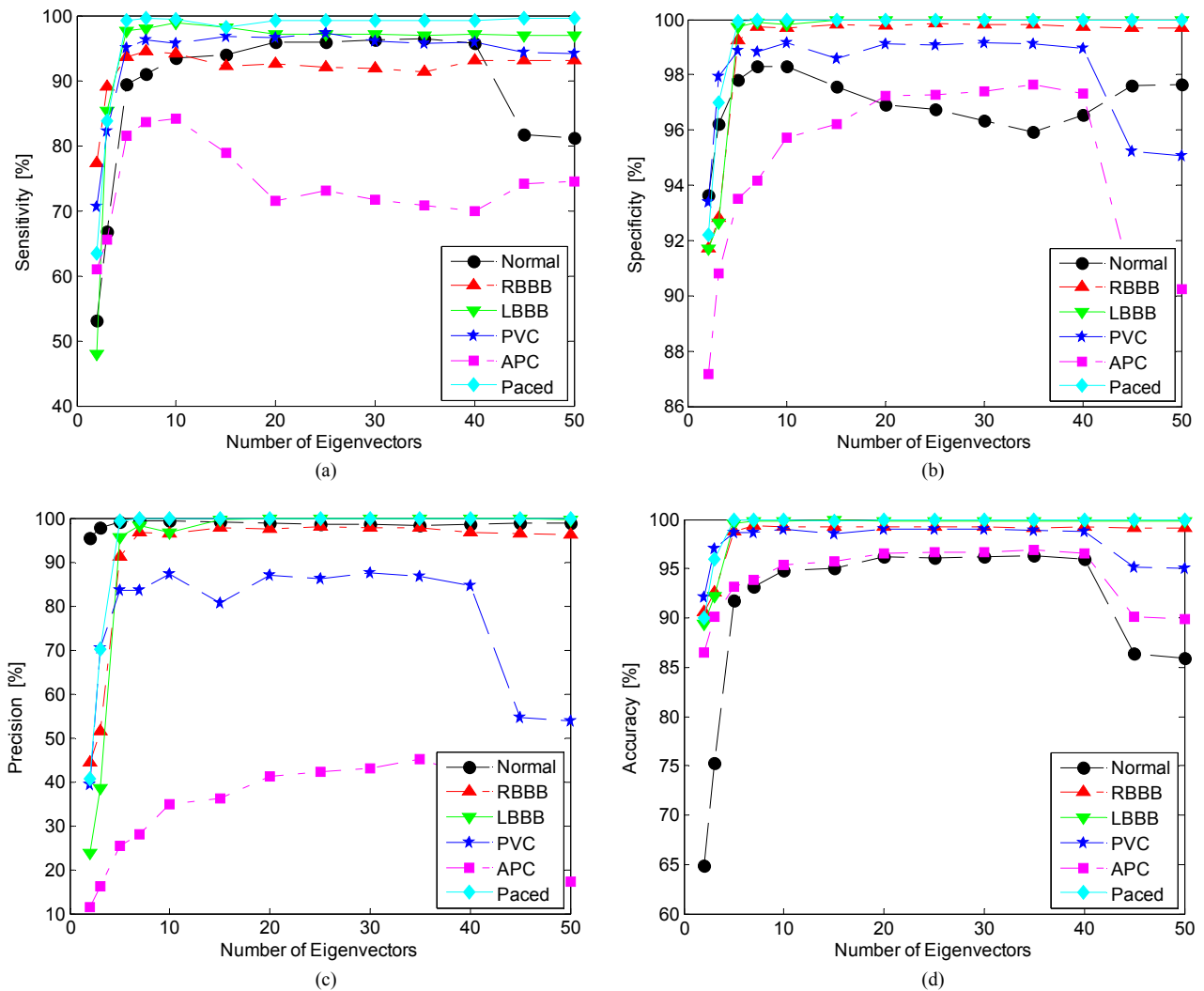


Figure 2. Classifier (a) sensitivity, (b) specificity, (c) precision, and (d) accuracy for different beat types as functions of the KLT components' number.

To compare the performance of the proposed technique to that of other methods, we run the same tests also with a simple-to-use but state-of-the-art SVM-based classifier, using the reference implementation [51], without any kind of signal preprocessing or feature extraction. The same training and testing sets were used to make the results directly comparable. The confusion matrix and performance indices of the SVM classifier are reported in Table VI and Table VII respectively.

As can be seen from Fig. 2 (a)–(d), the trends of all the performance indices are quite similar, showing a sharp increase as the number of eigenvectors M rises to about five, then it remains almost stable until, for very high values of M , there is a significant drop in performance. This is due to the fact that the number of observations available in the training set is not large enough to obtain an accurate statistical model of a high-dimensional PDF with the EM algorithm employed. In fact, the APC class performance is almost always well below that of other classes, as it is the class with

the least number of observations (as can be seen in Table I, APC annotations amount to about one half of the second least numerous class).

TABLE V. PERFORMANCE ANALYSIS OF THE PROPOSED ECG BEAT CLASSIFIER OBTAINED BY CONSIDERING $M = 10$ KLT COMPONENTS

Class	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
<i>Normal</i>	93.39	98.30	99.26	94.82
<i>RBBB</i>	94.07	99.71	96.49	99.26
<i>LBBB</i>	98.87	99.82	96.76	99.77
<i>PVC</i>	95.75	99.16	87.34	98.96
<i>APC</i>	84.06	95.71	34.83	95.40
<i>Paced</i>	99.43	99.98	99.80	99.94

TABLE VI. CONFUSION MATRIX FOR THE ECG BEAT CLASSIFICATION OBTAINED BY USING THE SVM CLASSIFIER

Input	Recognized					
	Normal	RBBB	LBBB	PVC	APC	Paced
Normal	31712	9	10	41	40	0
RBBB	12	3439	0	4	84	0
LBBB	8	0	2285	5	0	0
PVC	71	4	3	2485	0	1
APC	247	84	0	2	859	0
Paced	4	0	0	2	0	3491

TABLE VII. PERFORMANCE ANALYSIS OF THE SVM CLASSIFIER

Class	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
Normal	99.69	97.39	98.93	99.02
RBBB	97.17	99.77	97.26	99.56
LBBB	99.43	99.97	99.43	99.94
PVC	96.92	99.87	97.87	99.70
APC	72.06	99.72	87.39	98.98
Paced	99.83	100.00	99.97	99.98

Because of this, we deemed that $M=10$ eigenvectors were a good compromise given the amount of available training data, resulting in an overall accuracy (number of correctly classified beats over total number of tested beats) of 94.1%. In this particular case the number F of Gaussian mixture components used by the EM algorithm for modeling the PDFs of the considered classes is 19 for normal, 5 for RBBB, 4 for LBBB, 8 for VPC, 6 for APC and 5 for paced beats as can be easily seen in Table VIII. Here the number F of Gaussian mixture components used by the EM algorithm to model the PDFs of every beat type for different values of M ($M = 2, 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45,$ and 50) has been reported. Table VIII also reports, in the last column for different values of M , the total number of Gaussian mixture components $\sum_s F^{(s)}$ used for modeling all the six multivariate PDFs.

The overall accuracy was also computed, and graphed, against the number or eigenvectors M , and results are reported in Fig. 3. As it easy to see, it would have been possible to achieve a slightly better maximum overall accuracy of 95.5% using 35 eigenvectors, but at the cost of a sensibly worse specificity and sensitivity.

Fig. 3 also reports, as a reference benchmark, in a dashed red line, the overall accuracy of the SVM classifier, which is 98.6%. The SVM approach clearly is able of obtaining better

TABLE VIII. NUMBER OF GAUSSIAN MIXTURE COMPONENTS (F) USED BY THE EM ALGORITHM TO MODEL THE PDFS OF EVERY BEAT TYPE FOR DIFFERENT VALUES OF M (KLT COMPONENTS)

M	Beat Type						TOT.
	Normal	RBBB	LBBB	PVC	APC	Paced	
2	21	10	7	18	13	12	81
3	24	11	17	21	13	11	97
5	25	10	8	19	10	14	86
7	25	9	6	13	8	9	70
10	19	5	4	8	6	5	47
15	9	3	3	6	3	3	27
20	8	3	2	4	1	3	21
25	5	3	2	3	1	3	17
30	5	3	2	2	1	3	16
35	4	3	2	2	1	3	15
40	4	3	2	2	1	3	15
45	2	2	2	1	1	3	11
50	2	2	2	1	1	3	11

results, but it is a much more complicated algorithm with very high computational requirements. In our tests the training of the model with SVM required more than 30 hours of computation on a PC equipped with an Intel Q9550 CPU running at 2.83 GHz, while the proposed approach is able to train a model in a few minutes.

Moreover, the model complexity must also be taken into account. The SVM model required 3257 support vectors to achieve its performance, for a total of $K = 257318$ model parameters. This can be a symptom of overfitting, so we can expect real-world performance to be much lower for the SVM classifier. Our proposed technique, on the other hand, required only $K = \left(\sum_s F^{(s)}\right)(1 + M + M(M + 1)/2) + MN = 3822$ parameters, almost two orders of magnitude less.

A better comparison of the two models would make use of the Akaike information criterion (AIC), which in our case can be stated as [52]:

$$AIC = L \log(E/L) + 2K \tag{47}$$

where E is the residual sum of squares, so that it can be assumed $E/L = 1 - a$, a being the overall accuracy. Lower values for the AIC usually denote better models, as they are able to capture the data with less free parameters. With the above mentioned data, the SVM classifier yields an AIC of about $0.32 \cdot 10^6$, while the proposed GMM-based classifier yields an AIC of only $-0.12 \cdot 10^6$, showing the clear advantage of the proposed approach in terms of model compactness.

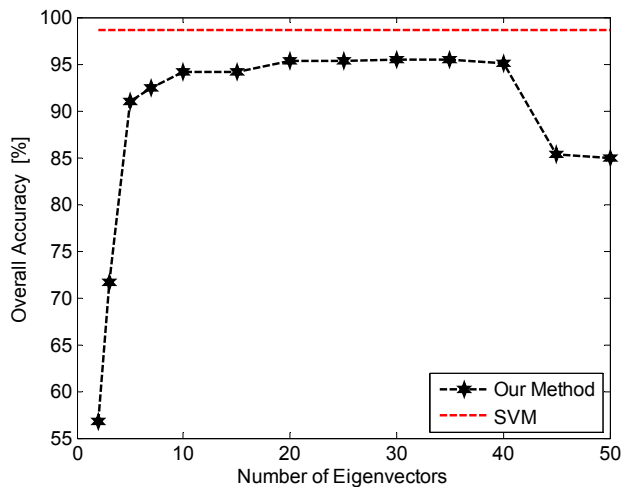


Figure 3. Classifier overall accuracy as a function of the KLT components' number.

VI. CONCLUSION

In this paper a novel technique suitable for a low-complexity multi-class ECG signal classification has been presented.

It is based on a statistical identification of nonlinear nonstationary signals. This methodology extracts the multivariate relationships of the ECG signal natural modes by exploiting the separability properties of the KLT. Then, by applying to the KLT-converted data an EM algorithm that is able to automatically find the optimal complexity, the parameters and the components of a Gaussian mixture model, it obtains the final statistical model maximizing the likelihood that the KLT-converted data are given by the estimated PDF. Based on the above statistical characterization, it is finally possible to perform a multi-class beat type classification.

Several simulations on ECG signals affected by arrhythmia were performed in order to test the validity of this technique. The experimental results show the very good performance of the classification technique considering the relative compactness of the employed model compared to that of other techniques, such as SVM. This is a clear advantage if the automatic detection needs to be performed on portable equipment with limited computational capabilities.

REFERENCES

- [1] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press, 2005.
- [2] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2008.
- [3] G. Biagetti, P. Crippa, A. Curzi, S. Orcioni, and C. Turchetti, "Analysis of the EMG signal during cyclic movements using multicomponent AM-FM decomposition," *IEEE Journal of Biomedical and Health Informatics*, (in press).
- [4] S. Mendis, P. Puska, and B. Norrving, *Global Atlas on cardiovascular disease prevention and control*. Geneva, Switzerland: World Health Organization, 2011.
- [5] H. Xia, I. Asif, and X. Zhao, "Cloud-ECG for real time ECG monitoring and analysis," *Computer Methods and Programs in Biomedicine*, vol. 110, no. 3, pp. 253–259, 2013.
- [6] Z. Zidelmal, A. Amirou, D. Ould-Abdeslam, and J. Merckle, "ECG beat classification using a cost sensitive classifier," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 3, pp. 570–577, 2013.
- [7] S. Conforto, A. Laudani, F. Oliva, F. Fulginei, and M. Schmid, "Classification of ECG patterns for diagnostic purposes by means of neural networks and support vector machines," in *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, 2013, pp. 591–595.
- [8] M. Llamedo and J. Martinez, "An automatic patient-adapted ECG heartbeat classifier allowing expert assistance," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2312–2320, 2012.
- [9] G. de Lannoy, D. Francois, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 241–247, 2012.
- [10] M. Bashir, D. G. Lee, M. Li, J.-W. Bae, H. S. Shon, M. C. Cho, and K.-H. Ryu, "Trigger learning and ECG parameter customization for remote cardiac clinical care information system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 561–571, 2012.
- [11] M. Llamedo and J. Martinez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 616–625, 2011.
- [12] I. Khalil and F. Sufi, "CardioGrid: ECG analysis on demand to detect cardiovascular abnormalities," in *9th International Conference on Information Technology and Applications in Biomedicine (ITAB 2009)*, 2009, pp. 1–5.
- [13] C. Meyer, J. Gavela, and M. Harris, "Combining algorithms in automatic detection of QRS complexes in ECG signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 468–475, 2006.
- [14] M. O'Dwyer, P. De Chazal, and R. Reilly, "Beat classification for use in arrhythmia analysis," in *Computers in Cardiology 2000*, 2000, pp. 395–398.
- [15] P. Melin, J. Amezcua, F. Valdez, and O. Castillo, "A new neural network model based on the LVQ algorithm for multi-class classification of arrhythmias," *Information Sciences*, vol. 279, pp. 483–497, 2014.
- [16] K. Daqrouq, A. Alkhateeb, M. Ajour, and A. Morfeq, "Neural network and wavelet average framing percentage energy for atrial fibrillation classification," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 3, pp. 919–926, 2014.
- [17] J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, and Y.-T. C. Yang, "ECG arrhythmia classification using a probabilistic neural network with a feature reduction method," *Neurocomputing*, vol. 116, pp. 38–45, 2013.
- [18] R. Kumar and Y. Kumaraswamy, "Investigation and classification of ECG beat using input output additional weighted feed forward neural network," in *2013 International Conference on Signal Processing Image Processing Pattern Recognition (ICSIPR)*, vol. 1, 2013, pp. 200–205.
- [19] S. Faziludeen and P. Sabiq, "ECG beat classification using wavelets and SVM," in *2013 IEEE Conference on Information Communication Technologies (ICT)*, 2013, pp. 815–818.
- [20] J. Abawajy, A. Kelarev, and M. Chowdhury, "Multistage approach for clustering and classification of ECG data," *Computer Methods and Programs in Biomedicine*, vol. 112, no. 3, pp. 720–730, 2013.

- [21] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomedical Signal Processing and Control*, vol. 8, no. 5, pp. 437–448, 2013.
- [22] M. Kallas, C. Francis, L. Kanaan, D. Merheb, P. Honeine, and H. Amoud, "Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals," in 2012 19th International Conference on Telecommunications (ICT), 2012, pp. 1–5.
- [23] C. Ye, B. Kumar, and M. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930–2941, 2012.
- [24] M. Nait-Hamoud and A. Moussaoui, "Two novel methods for multiclass ECG arrhythmias classification based on PCA, fuzzy support vector machine and unbalanced clustering," in 2010 International Conference on Machine and Web Intelligence (ICMWI), 2010, pp. 140–145.
- [25] P. De Chazal, M. O'Dwyer, and R. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
- [26] P. De Chazal and R. Reilly, "Automatic classification of ECG beats using waveform shape and heart beat interval features," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03)*, vol. 2, 2003, pp. II–269–272.
- [27] I. Saini, D. Singh, and A. Khosla, "Electrocardiogram beat classification using empirical mode decomposition and multiclass directed acyclic graph support vector machine," *Computers & Electrical Engineering*, vol. 40, no. 5, pp. 1774–1787, 2014.
- [28] Z. Dokur and T. Ölmez, "ECG beat classification by a novel hybrid neural network," *Computer Methods and Programs in Biomedicine*, vol. 66, no. 2-3, pp. 167–181, 2001.
- [29] A. E. Zadeh, A. Khazaei, and V. Ranaee, "Classification of the electrocardiogram signals using supervised classifiers and efficient features," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 2, pp. 179–194, 2010.
- [30] E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of ECG signals using Lyapunov exponents," *Computer Methods and Programs in Biomedicine*, vol. 93, no. 3, pp. 313–321, 2009.
- [31] Y. P. Meau, F. Ibrahim, S. A. Narainasamy, and R. Omar, "Intelligent classification of electrocardiogram (ECG) signal using extended Kalman Filter (EKF) based neuro fuzzy system," *Computer Methods and Programs in Biomedicine*, vol. 82, no. 2, pp. 157–168, 2006.
- [32] M. Engin, "ECG beat classification using neuro-fuzzy network," *Pattern Recognition Letters*, vol. 25, no. 15, pp. 1715–1722, 2004.
- [33] O. Sayadi and M. Shamsollahi, "Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a Bayesian filter," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2748–2757, Oct 2011.
- [34] O. Sayadi, M. Shamsollahi, and G. Clifford, "Robust detection of premature ventricular contractions using a wave-based Bayesian framework," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 353–362, Feb 2010.
- [35] R. Andreao, B. Dorizzi, and J. Boudy, "ECG signal analysis through hidden Markov models," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, pp. 1541–1549, Aug 2006.
- [36] G. Biagetti, P. Crippa, A. Curzi, S. Orcioni, and C. Turchetti, "A multi-class ECG beat classifier based on the truncated KLT representation," in 2014 UKSim-AMSS 8th European Modelling Symposium on Mathematical Modelling and Computer Simulation, (EMS2014), October 2014, pp. 93–98.
- [37] F. Gianfelici, C. Turchetti, and P. Crippa, "A non-probabilistic recognizer of stochastic signals based on KLT," *Signal Processing*, vol. 89, no. 4, pp. 422–437, 2009.
- [38] M. A. F. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, Mar 2002.
- [39] H. Yin and N. Allinson, "Self-organizing mixture networks for probability density estimation," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 405–411, Mar. 2001.
- [40] P. Crippa, F. Gianfelici, and C. Turchetti, "Information theoretical algorithm based on statistical models for blind identification of nonstationary dynamical systems," in *International Joint Conference on Neural Networks, (IJCNN 2009)*, June 2009, pp. 612–618.
- [41] G. Biagetti, P. Crippa, A. Curzi, and C. Turchetti, "Unsupervised identification of nonstationary dynamical systems using a Gaussian mixture model based on EM clustering of SOMs," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, June 2010, pp. 3509–3512.
- [42] R. Mark and G. Moody, "MIT-BIH arrhythmia database," May 1997.
- [43] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, 2000.
- [44] G. Prasad and J. S. Sahambi, "Classification of ECG arrhythmias using multi-resolution analysis and neural networks," in *Conference on Convergent Technologies for the Asia-Pacific Region (TENCON 2003)*, vol. 1, 2003, pp. 227–231.
- [45] 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, pp. 582–589, April 2004.
- [46] P. De Chazal and R. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2535–2543, 2006.
- [47] Y.-H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 891–900, 1997.
- [48] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1992.
- [49] T. T. Soong, *Random Differential Equations in Science and Engineering*. New York, NJ, USA: Academic Press, 1973.
- [50] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, Jul 2000.
- [51] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.
- G. Panchal, A. Ganatra, Y. Kosta, and D. Panchal, "Searching most efficient neural network architecture using Akaike's information criterion (AIC)," *International Journal of Computer Applications*, vol. 1, no. 5, pp. 41–44, 2010.