



A wavelet optimization approach for ECG signal classification

Abdelhamid Daamouche^a, Latifa Hamami^b, Naif Alajlan^c, Farid Melgani^{a,*}

^a Department of Information and Communication Technologies, University of Trento, Trento, Italy

^b School of Engineering, El Harrach, Algiers, Algeria

^c ALISR Laboratory, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

ARTICLE INFO

Article history:

Received 5 May 2011

Received in revised form 14 June 2011

Accepted 4 July 2011

Available online 30 July 2011

Keywords:

Classification

Discrete wavelet transform (DWT)

Electrocardiogram (ECG) signals

Particle swarm optimization (PSO)

Support vector machines (SVM)

ABSTRACT

Wavelets have proved particularly effective for extracting discriminative features in ECG signal classification. In this paper, we show that wavelet performances in terms of classification accuracy can be pushed further by customizing them for the considered classification task. A novel approach for generating the wavelet that best represents the ECG beats in terms of discrimination capability is proposed. It makes use of the polyphase representation of the wavelet filter bank and formulates the design problem within a particle swarm optimization (PSO) framework. Experimental results conducted on the benchmark MIT/BIH arrhythmia database with the state-of-the-art support vector machine (SVM) classifier confirm the superiority in terms of classification accuracy and stability of the proposed method over standard wavelets (i.e., Daubechies and Symlet wavelets).

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The electrocardiogram (ECG) signal represents the changes in electrical potential during the cardiac cycle as recorded between surface electrodes on the body [1]. The analysis of ECG signals can provide clinicians with valuable information about the patient health condition. In this context, significant research efforts have been devoted for developing automatic and fast arrhythmia diagnosis tools based on the processing and analysis of ECG signals.

In the last two decades, wavelets have attracted a growing interest in many signal processing and analysis applications. The main interesting feature of wavelets is their time-frequency representation of the signal. They allow gaining a deep insight of the signal at different scales and frequencies, and have proved particularly successful both in ECG signal compression and classification [1–13].

In the context of ECG signal classification which represents the focus of this paper, several interesting works can be found in the literature. In particular, in [4], Ince et al. proposed a feature extraction technique that employs the translation-invariant dyadic wavelet transform in order to effectively extract the morphological information from ECG data. In [5], Sahambi et al. presented an approach

that uses a dyadic wavelet to characterize the ECG signal. To circumvent its high computational cost, they used digital signal processing add-on cards. In [6], a method for detecting premature ventricular contraction (PVC) from the Holter system is proposed using wavelet transform and fuzzy neural network. In [7], Dickhaus et al. addressed two questions: how are the recorded time courses of the signals to be interpreted with regard to a diagnostic decision? What are the essential features and how is the information hidden in the signals? Then they presented an example to identify patients who are at high-risk of developing ventricular tachycardia (VT). In [8], an approach to detect PVCs using a neural network with weighted fuzzy membership functions is described. To discriminate between normal and PVC beats, Lim et al. exploited wavelet coefficients. In [9], a dyadic wavelet transform is used for extracting ECG characteristic points. The local maxima of the wavelet modulus at different scales are used to locate the sharp variation points of ECG. The proposed algorithm first detects the QRS complex, then the T wave, and finally the P wave. In [10], Khamene and Negahdaripour proposed a solution that relies on the positions of singular points (high peaks) of the ECG signal. Their method attempts to discriminate between the singular points of the maternal and fetal ECGs, both present in the composite abdominal signal. All the work is carried out in the wavelet transformed space of the ECG signal. In [11], Inan et al. presented an approach for classifying beats of a large dataset by training a neural network classifier using wavelet and timing features. They found that the fourth scale of a dyadic wavelet transform with a quadratic spline wavelet together with

* Corresponding author.

E-mail address: melgani@disi.unitn.it (F. Melgani).

the pre/post RR-interval ratio is very effective in distinguishing normal and PVC from other beats. In [12], features extracted from successive wavelet coefficient levels after wavelet decomposition of signals of heart rate variability (HRV) from RR intervals and ECG-derived respiration (EDR) from R waves of QRS amplitudes were used as inputs to a support vector machine (SVM) classifier to recognize obstructive sleep apnea syndrome. In [13], Senhaji et al. raised an important question: what is the most appropriate wavelet to use? The answer was: there is no theoretical answer at the moment and the choice must be done empirically by comparing results of different wavelets.

All the aforementioned works made use of wavelets which have been derived for general signal processing and analysis. However, we believe that in order to improve wavelet performances in ECG classification, one should design wavelets that are optimized for this specific problem. This paper is intended to propose a wavelet design method which is driven by the classification process performance in terms of accuracy. Due to the very complex relationship characterizing the wavelet and the classifier accuracy, we resort to a stochastic design method based on particle swarm optimization (PSO) which has proved capable to provide effective answers to problems raised by various applications [14–16]. The proposed method exploits the polyphase representation of the discrete wavelet transform (DWT). Such representation allows generating a wavelet filter bank from a set of angular parameters, and thus formulating the wavelet design problem for ECG signal classification as a problem of estimating these parameters so that to maximize the classifier accuracy. The kind of classification approach adopted in this work is the state-of-the-art SVM classifier known for its high generalization capability.

The remaining of this paper is organized as follows. Section 2 gives a general review of wavelets. Sections 3 and 4 present the main principles of PSO and SVM, respectively. Section 5 describes the proposed wavelet design method. Experimental results are provided in Section 6. Finally, conclusions are drawn in Section 7.

2. Wavelets

The wavelet transform is a linear operation that decomposes a signal into components that appear at different scales [1,5,17]. Wavelet functions $\Psi(t)$ are defined in a space of measurable functions that are absolute and square integrable, i.e.,

$$\int_{-\infty}^{+\infty} |\Psi(t)| dt < \infty \quad (1)$$

$$\int_{-\infty}^{+\infty} |\Psi(t)|^2 dt < \infty \quad (2)$$

In such a space, they should satisfy conditions of zero mean and square norm one [17]:

$$\int_{-\infty}^{+\infty} \Psi(t) dt = 0 \quad (3)$$

$$\int_{-\infty}^{+\infty} |\Psi(t)|^2 dt = 1 \quad (4)$$

The wavelet transform of a function $f(t) \in L^2(R)$ at scale a and position τ is given by [5]:

$$Wf(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \Psi^* \left(\frac{t - \tau}{a} \right) dt \quad (5)$$

The asterisk $*$ denotes the complex conjugation.

Eq. (5) means that the signal to be analyzed $f(t)$ is convolved with stretched/dilated copies of the mother wavelet $\Psi(t)$. For $a < 1$, the wavelet is contracted and the transform gives information about the finer details of $f(t)$. For $a > 1$, the wavelet expands and the transform gives a coarse view of the signal. If the scale parameter $a = 2^j$ with $j \in \mathbb{Z}$, \mathbb{Z} is an integer set, then the wavelet is called a dyadic wavelet [17]. The wavelet transform operates in continuous time on functions and in discrete time on vectors. In continuous time, the wavelet coefficients are found by evaluating the integral in (5). Whereas, in discrete time, the coefficients are found by passing a vector ($x(n)$, n integer) through a bank of two filters, one is a low-pass and the other is a high-pass.

A complete and interesting characterization of the DWT filter coefficients with compact support was presented by Daubechies in [18]. However, in general, since looking for an optimum wavelet is a problem-dependent issue, DWT design can take many forms. In this context, an elegant way to determine the coefficients of a filter bank has been developed by Sherlock and Monro [19]. It is a polyphase method [20] which relies on a factorization proposed by Vaidyanathan [21]. Their algorithm allows deriving any orthonormal perfect-reconstruction finite impulse response (FIR) filter of arbitrary length. In the following, the method is briefly described. The low-pass filter coefficients in the z -domain are given by:

$$H_0(z) = \sum_{i=0}^{2N-1} h_i z^{-i} \quad (6)$$

and thus

$$H_0(z) = \sum_{i=0}^{N-1} h_{2i} z^{-2i} + z^{-1} \sum_{i=0}^{N-1} h_{2i+1} z^{-2i} \quad (7)$$

In (7), H_0 is decomposed into even and odd powers of z . Vaidyanathan proposed the following factorization of the polyphase matrix [21]:

$$H_p(z) = \begin{pmatrix} H_{00}(z) & H_{01}(z) \\ H_{10}(z) & H_{11}(z) \end{pmatrix} = \begin{pmatrix} c_0 & s_0 \\ -s_0 & c_0 \end{pmatrix} \prod_{i=1}^{N-1} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \begin{pmatrix} c_i & s_i \\ -s_i & c_i \end{pmatrix} \quad (8)$$

where

$$H_{00}(z) = \sum_{i=0}^{N-1} h_{2i} z^{-2i} \quad (9)$$

and

$$H_{01}(z) = \sum_{i=0}^{N-1} h_{2i+1} z^{-2i} \quad (10)$$

$H_{00}(z)$ and $H_{01}(z)$ represent the polyphase components of the low-pass filter, whereas $H_{10}(z)$ and $H_{11}(z)$ are those of the high-pass filter. The coefficients c_i and s_i are computed as follows: $c_i = \cos(\theta_i)$ and $s_i = \sin(\theta_i)$.

Sherlock and Monro developed a new formulation by rewriting the factorization in a recursive form [19]:

$$H_p^{(k+1)}(z) = H_p^{(k)}(z) \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \begin{pmatrix} c_k & s_k \\ -s_k & c_k \end{pmatrix} \quad (11)$$

with $k = 1, 2, \dots, N$ and $H_p^{(a)} = \begin{pmatrix} c_0 & s_0 \\ -s_0 & c_0 \end{pmatrix}$. The superscript (k) refers to filters of length $2k$. This form leads to the following recursive formulae for the even-numbered filter coefficients:

$$\begin{cases} h_0^{(k+1)} = c_k h_0^{(k)} \\ h_{2i}^{(k+1)} = c_k h_{2i}^{(k)} - s_k h_{2i-1}^{(k)} \quad \text{for } i = 1, 2, \dots, k-1 \\ h_{2k}^{(k+1)} = -s_k h_{2k-1}^{(k)} \end{cases} \quad (12)$$

with $h_0^{(1)} = c_0$ and $h_1^{(1)} = s_0$.

The formulae for the odd coefficients are given by:

$$\begin{cases} h_1^{(k+1)} = s_k h_1^{(k)} \\ h_{2i+1}^{(k+1)} = s_k h_{2i}^{(k)} + c_k h_{2i-1}^{(k)} \quad \text{for } i = 1, 2, \dots, k-1 \\ h_{2k+1}^{(k+1)} = c_k h_{2k}^{(k)} \end{cases} \quad (13)$$

Eqs. (12) and (13) express the low-pass coefficients $\{h_0, h_1, \dots, h_{2N-1}\}$ in terms of N free chosen angular parameters $\{\theta_0, \theta_1, \dots, \theta_{N-1}\}$ whose values are in the interval $[0, 2\pi)$. The high-pass filter coefficients are found by alternating flip construction, that is

$$g_i = (-1)^{i+1} h_{2N-1-i} \quad (14)$$

From N free parameters, one can generate the $2N$ low-pass filter coefficients and the corresponding high-pass filter coefficients. Therefore, the design of an optimal DWT can be viewed as an optimization problem in the \mathcal{R}^N space of the angular parameters θ_i 's.

3. Particle swarm optimization

Particle swarm optimization (PSO) is a stochastic search method which was first proposed in 1995 by Kennedy and Eberhart [22]. PSO has proved promising in solving different pattern recognition problems [4,14–16]. Similar to evolutionary computation algorithms, such as genetic algorithms [23], PSO is a population-based optimization method which exhibits the advantages of implementation simplicity and few free parameters to adjust. It is inspired from swarm intelligence observed in bird flocking and fish schooling. The optimization method is based on sharing information among the swarm members called particles, so that each particle adjusts its velocity and direction according to its past and current trajectories and to the best position found among all the particles of the swarm. The position of each particle corresponds to a particular candidate of the solution space and thus to a specific value of the fitness function to be optimized.

For a problem with d real variables q_1, q_2, \dots and q_d to estimate, the algorithm starts from a random swarm of particles called initial population. Assume that the swarm is of size S . Each particle \mathbf{p}_i ($i = 1, 2, \dots, S$) from the swarm is characterized by: (a) a current position $\mathbf{p}_i(t) \in \mathcal{R}^d$, which refers to a candidate solution of the optimization problem at iteration t , i.e., $\mathbf{p}_i(t) = [q_{i1}(t), q_{i2}(t), \dots, q_{id}(t)]$; (b) a velocity $\mathbf{v}_i(t) \in \mathcal{R}^d$, and (c) a best position $\mathbf{p}_{bi}(t) \in \mathcal{R}^d$ which is found during its past trajectory ("best" with respect to the objective/fitness function to optimize). Let $\mathbf{p}_g(t) \in \mathcal{R}^d$ be the best global position found over all trajectories that were traveled by the particles of the swarm. In other words, the best global position corresponds to the best value of the fitness function ever found by the swarm up to instant t . During the search process, the particles move according to the following equations [22]:

$$\mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + c_1 \cdot \mathbf{r}_1(t)(\mathbf{p}_{bi}(t) - \mathbf{p}_i(t)) + c_2 \cdot \mathbf{r}_2(t)(\mathbf{p}_g(t) - \mathbf{p}_i(t)) \quad (15)$$

$$\mathbf{p}_i(t+1) = \mathbf{p}_i(t) + \mathbf{v}_i(t) \quad (16)$$

where $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ are random variables that are drawn from a uniform distribution in the range $[0,1]$ to provide a stochastic weighting for components involved in (15). The constants c_1 and c_2 regulate the relative velocities with respect to the best local and global positions, respectively. The inertia weight w is used as a tradeoff between global and local exploration capabilities of the swarm. Eqs. (15) and (16) are iterated until convergence is reached. Note that the particle velocity just expresses a move since time unit is assumed to be equal to one and, therefore, like the particle position, it also encodes the d real variables q_1, q_2, \dots and q_d to estimate. In PSO, the information sharing mechanism is of the one-way type since only $\mathbf{p}_g(t)$ gives out the information to other particles.

Typical convergence criteria are based on the iterative behavior of the best value of the adopted fitness function or/and simply on a user-defined maximum number of iterations.

4. Support vector machines

As a classification approach, we will adopt in this work the state-of-the-art support vector machine (SVM) approach which has shown particularly effective in numerous application fields including the classification of ECG signals [16,24,25]. The focus on the SVM classifier is motivated by its commonly admitted superiority over traditional classifiers. Note that the method proposed in this paper is general and, therefore, any other kind of classifier could be considered as well. In the following, we will briefly introduce this tool. For further details, we refer the reader to [26].

Let us first consider for simplicity a supervised binary classification problem. Let us assume that the training set consists of S vectors $\mathbf{x}_i \in \mathcal{R}^d$ ($i = 1, 2, \dots, S$) from the d -dimensional feature space \mathbf{X} . The d features are for instance morphological and timing features extracted from the ECG signal. To each vector \mathbf{x}_i , we associate a target $y_i \in \{-1, +1\}$ (e.g., normal and abnormal beats, respectively). The linear SVM classification approach consists of looking for a separation between the two classes in \mathbf{X} by means of an optimal hyperplane that maximizes the separating margin [26]. In the non-linear case, which is the most common case since data are often not linearly separable, the two classes are first mapped with a kernel method in a higher dimensional feature space, i.e., $\Phi(\mathbf{X}) \in \mathcal{R}^{d'}$ ($d' > d$). The membership decision rule is based on the function $\text{sign}[f(\mathbf{x})]$, where $f(\mathbf{x})$ represents the discriminant function associated with the hyperplane in the transformed space and is defined as:

$$f(\mathbf{x}) = \mathbf{w}^* \cdot \Phi(\mathbf{x}) + b^* \quad (17)$$

The optimal hyperplane defined by the weight vector $\mathbf{w}^* \in \mathcal{R}^{d'}$ and the bias $b^* \in \mathcal{R}$ is the one that minimizes a cost function that expresses a combination of two criteria: margin maximization and error minimization [26]. The solution of such optimization problem is a discriminant function conveniently expressed as a function of the data in the original (lower) dimensional feature space \mathbf{X} :

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \quad (18)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$ is the vector of Lagrange multipliers and $K(\cdot, \cdot)$ is a kernel function. The set S is a subset of the indices $\{1, 2, \dots, S\}$ corresponding to the non-zero Lagrange multipliers α_i^* 's, which define the so-called support vectors.

The kernel $K(\cdot, \cdot)$ must satisfy the condition stated in Mercer's theorem so as to correspond to some type of inner product in the

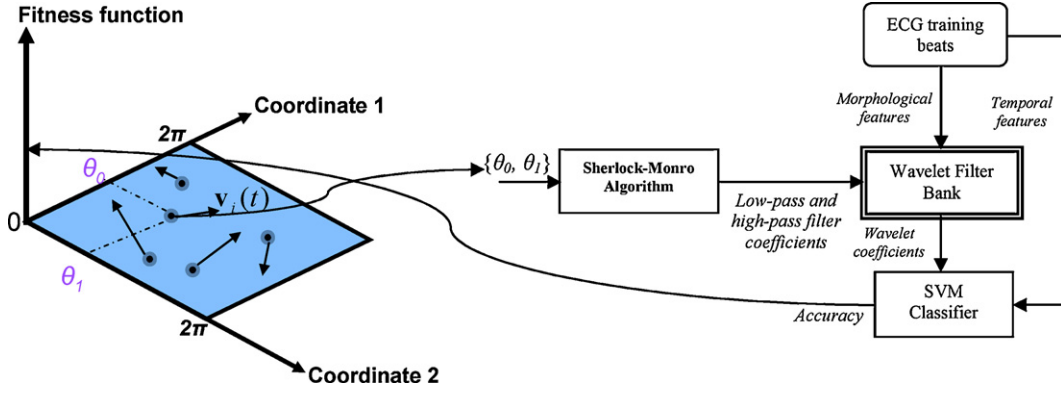


Fig. 1. Illustration of the PSO search space for second-order filters (two angular parameters to be estimated) and its relationship with the wavelet filter design.

transformed (higher) dimensional feature space $\Phi(\mathbf{X})$ [25]. A typical example of such kernels is represented by the well-known Gaussian function:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp \left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2 \right) \quad (19)$$

where γ represents a parameter inversely proportional to the width of the Gaussian kernel.

As described above, SVMs are intrinsically binary classifiers. But the classification of ECG signals often involves the simultaneous discrimination of several information classes (arrhythmias). In order to face this issue, a number of multiclass classification strategies can be adopted [27]. In this paper, we considered the popular one-against-one (OAO) strategy. Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_T\}$ be the set of T possible classes. The idea behind this strategy is to train an ensemble of binary SVMs, each for any two classes, and to construct a global decision by exploiting the partial ones yielded by the single SVMs. Thus, for T classes, the OAO strategy involves $T(T-1)/2$ SVMs, each representing a discriminant function f_{ij} between classes ω_i and ω_j .

5. The proposed method

Although wavelets have been developed for general signal processing purposes, as mentioned in Section 1, the literature reports many works exploiting this tool for analyzing ECG signals. Most of the existing techniques are based on the well-known Daubechies wavelet. However, we believe that the wavelet performance could be improved if the design is driven by the ECG classification accuracy. Therefore, we propose a wavelet design method which adopts the polyphase representation and formulates the optimization problem within a PSO framework. As shown in Fig. 1, the coordinates of the particles of the swarm encode the angular parameters $\{\theta_0, \theta_1, \dots, \theta_{N-1}\}$.

Concerning the fitness function, we use the classification accuracy of the state-of-the-art SVM classifier achieved by cross-validation (CV) on the training set [28]. This fitness function is an empirical estimate of the generalization accuracy that drives the PSO during the search process toward the best wavelet candidate in the wavelet space. In greater detail, during the training phase, the SVM parameters (regularization and kernel parameters) are selected according to an m -fold CV procedure (typically, m is chosen between 2 and 10), first by randomly splitting the S available training beats into m mutually exclusive subsets (folds) of equal size, and then by training m times an SVM classifier modeled with predefined parameter values. Each time, one of the subsets is left out of the training and is only used (as a set of validation sam-

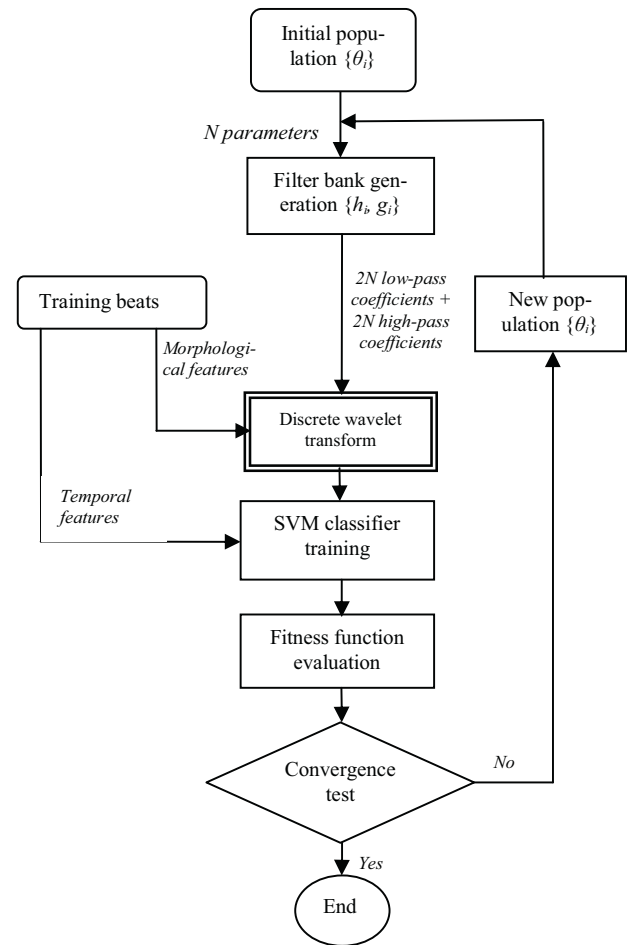


Fig. 2. Block diagram of the PSO search process.

ples) to obtain an estimate of the classification accuracy, while the remaining $m-1$ folds are used to train the classifier. From m times of training and accuracy computation, the average accuracy yields a prediction of the classification accuracy of the considered SVM classifier. We select the best SVM classifier parameter values to maximize this prediction and get the final classifier accuracy estimate. The leave-one-out (LOO) procedure is another empirical procedure which could be adopted for getting such estimate. It is generally more accurate but also much more computationally demanding since it is equivalent to an S -fold CV.

As illustrated in Fig. 2, the main steps of the proposed wavelet design method are described in the following:

Phase 1: Initialization

- Set filter order $N = \text{order_min}$.
- Set decomposition level $L = \text{decLev_min}$

Phase 2: Starting population

- Initialize the PSO with a random population by generating for each particle a position vector $(\{\theta_0, \theta_1, \dots, \theta_{N-1}\})$ of random values uniformly distributed in the range $[0, 2\pi)$. Set the velocity vector of each particle to zero.
- For each particle, compute the corresponding fitness function value through the following steps:
 - inject the vector of θ_i 's into the recursive algorithm (12) and (13) in order to generate the 2N low-pass filter coefficients;
 - by alternating flip construction, derive the 2N high-pass filter coefficients (14);
 - with the resulting low-pass and high-pass filters, apply DWT to each ECG training beat;
 - train an SVM classifier by feeding it with the generated wavelet features (and, if desired, other feature types). Compute its cross-validation accuracy $(CVA_i(t))$ to set the fitness function value of the particle, $p_i(0) = CVA_i(0)$ for $i = 0, 1, \dots, N - 1$.
- Store the position of each particle and label it as best local position. Save the position of the particle with the largest fitness function value as the best global position.

$$p_{bi}(0) = p_i(0) \text{ and } p_g(0) = \max(p_{bi}(0))$$

Phase 3: Search process

- Update the velocity vector of each particle using (15).
- Update the particle coordinates according to (16).
- For each particle, compute again the fitness function value (cross-validation accuracy) in the corresponding new wavelet domain, $p_i(t) = CVA_i(t)$.
- Update the best local position $p_{bi}(t)$ for each particle and the best global position, $p_g(t) = \max(p_{bi}(t))$.

Phase 4: Convergence check

If the number of generations is different from the user-defined maximum number of generations, or if the best fitness function value (best global position), still varies significantly in the last iterations, i.e., $|p_g(t) - p_g(t-1)| > \varepsilon$ (where ε is a user-defined threshold), go to Phase 3, otherwise end the search.

Phase 5: Filter order and decomposition level

- For each filter order and for each decomposition level in the predefined ranges $[\text{order_min}, \text{order_max}]$ and $[\text{decLev_min}, \text{decLev_max}]$, respectively, with a predefined moving pace Δ (e.g., $\Delta = 1$), go through Phases 2–4.
- Select the couple of values (of filter order and decomposition level) which yields the highest fitness function value at convergence.

6. Experimental results

6.1. Dataset description and experimental setup

Our experiments were conducted on the basis of ECG data from the MIT-BIH arrhythmia database [29]. In particular, the considered beats refer to the following classes: normal sinus rhythm (denoted in the following as 'N'), atrial premature beat ('A', irregular beat which starts in the atria, i.e., the upper two chambers of the heart), ventricular premature beat ('V', beat initiated by the heart ventricles rather than by the sinoatrial node), right bundle branch block ('RB', causes prolongation of the last part of the QRS complex and may shift the heart electrical axis to the right), left bundle branch block ('LB', widens the entire QRS and typically shifts the heart electrical axis to the left), and paced beat ('/'). The beats were selected from the recordings of 20 patients representatives of these classes, which correspond to the following files: 100, 102, 104, 105, 106, 107, 118, 119, 200, 201, 202, 203, 205, 208, 209, 212, 213, 214, 215, and 217. In order to feed the classification process, in this study we adopted the two following kinds of features: (i) ECG morphology features and (ii) three ECG temporal features, i.e. the QRS complex duration, the RR interval (the time span between two consecutive R points representing the distance between the QRS peaks of the present and previous beats), and the RR interval averaged over the ten last beats [30]. In order to extract these features, first we performed the QRS detection and ECG wave boundary recognition tasks by means of the well-known *ecgpuwave* software available

on <http://www.physionet.org/physiotools/ecgpuwave/src/>. Then after extracting the three temporal features of interest, we normalized to the same periodic length the duration of the segmented ECG cycles according to the procedure reported in [31]. To this purpose, the mean beat period was chosen as the normalized periodic length, which was represented by 300 uniformly distributed samples. Consequently, the total number of morphology and temporal features equals 303 for each beat.

The training set used in all experiments consists of only 125 samples selected randomly from the benchmark MIT-BIH arrhythmia database. The choice of this relatively small number of training samples (less than 1% of the total test set) is motivated by two reasons: (1) it reduces sharply the processing time required for convergence and (2) it permits to test the method in delicate situations where the number of available training beats is very limited. The detailed numbers of training beats and test beats are reported for each class in Table 1.

Classification performance was evaluated in terms of four measures, which are: (1) the overall accuracy (OA), which is the percentage of correctly classified beats among all the beats considered (independently of the classes they belong to); (2) the accuracy of each class that is the percentage of correctly classified beats among the beats of the considered class; (3) the average accuracy (AA), which is the average over the classification accuracies obtained for the different classes; and (4) the McNemar's test which gives the statistical significance of differences between the accuracies achieved by the different classification approaches. This test is based on the standardized normal test statistic [32]:

$$Z_{ij} = \frac{f_{ij} - f_{ji}}{\sqrt{f_{ij} + f_{ji}}} \quad (20)$$

where Z_{ij} measures the pairwise statistical significance of the difference between the accuracies of the i th and j th classifiers. f_{ij} stands for the number of beats classified correctly and wrongly by the i th and j th classifiers, respectively. Accordingly, f_{ij} and f_{ji} are the counts of classified beats on which the considered i th and j th classifiers disagree. At the commonly used 5% level of significance, the difference of accuracies between the i th and j th classifiers is said statistically significant if $|Z_{ij}| > 1.96$.

Moreover, we assessed the capability of the proposed wavelet design procedure to discriminate the ventricular premature beats (V) from the others. The immediate detection and treatment of the 'V' arrhythmia is considered very important since it can be linked to mortality when associated with myocardial infarction. In particular, we used three common measures [33]: (1) sensitivity Se ($Se = TP/(TP + FN)$), it is the ratio between the number of 'V' beats correctly detected and the total number of 'V' beats; (2) specificity (Sp) ($Sp = TN/(TN + FP)$), it stands for the fraction of 'non-V' beats that has been correctly rejected; and (3) positive predictivity Pp ($Pp = TP/(TP + FP)$), it defines the ratio of 'V' beats correctly detected over the total number of detected 'V' beats. In the above definitions, the notations are as follows: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

The PSO configuration is as follows: the size of the swarm is fixed to 10 and the maximum number of iterations to 20. The inertia weight w was set to 0.75, and $c_1 = c_2 = 1$. An example of behavior of the PSO during the search process is shown in Fig. 3, which illustrates the gradual convergence of the fitness function value associated with the best global swarm position. In all experiments reported in this paper, we adopted the SVM classifier with Gaussian kernel and a 5-fold CV. Finally, note that the proposed wavelet design method was applied on the 300 morphology features while the three temporal features were added after the wavelet decomposition with the filter encoded by the considered particle.

From previous works [4,11], it emerges that the best ECG signal wavelet decomposition is achieved up to the fourth decomposition

Table 1

Numbers of training and test beats used in the experiments. The classes are: normal sinus rhythm (N), atrial premature beat (A), ventricular premature beat (V), right bundle branch block (RB), left bundle branch block (LB), and paced beat (/).

Class	N	A	V	RB	/	LB	Total
Training beats	37	24	25	13	13	13	125
Test beats	24,000	238	3939	3739	6771	1751	40,438

Table 2

Overall accuracies achieved on the test beats by feeding the SVM classifier with wavelet features generated by the proposed wavelet (PSO-based), the Daubechies and the Symlet wavelets, with varying decomposition levels and filter orders. The best result for each wavelet is in boldface.

Filter order	PSO-based Decomposition level				Daubechies Decomposition level				Symlet Decomposition level			
	1	2	3	4	1	2	3	4	1	2	3	4
4	86.36	86.63	87.16	88.48	78.08	81.50	85.39	86.04	81.76	83.02	84.44	86.21
6	86.69	88.23	87.63	87.66	78.45	83.30	84.18	86.19	81.82	83.19	85.60	85.95
8	86.63	87.29	87.83	88.81	81.89	83.36	85.93	86.25	78.08	83.64	85.78	86.12
10	86.88	87.72	88.84	88.44	82.27	82.50	85.68	86.66	78.08	84.19	84.04	86.83

Table 3

Overall (OA), average (AA), and class percentage accuracies achieved on the test beats with the three wavelets. Only the best result is shown for each wavelet, the classes are: normal sinus rhythm (N), atrial premature beat (A), ventricular premature beat (V), right bundle branch block (RB), left bundle branch block (LB), and paced beat (/).

	Accuracy [%]							
	OA	AA	N	A	V	RB	/	LB
PSO-based	88.84	89.18	86.28	82.77	91.75	89.38	96.19	88.75
Daubechies	86.66	87.62	84.14	84.03	89.72	90.13	92.33	85.38
Symlet	86.83	86.50	84.67	81.09	85.40	90.42	94.55	82.87

The best classification results are shown in boldface.

level. Basing on this, in the following, we performed the performance assessment by considering the first four decomposition levels.

6.2. Results

Table 2 shows the accuracies obtained on the test beats by feeding the SVM classifier with the features generated by the proposed wavelet design method (PSO-based), and for comparison, by the Daubechies wavelet (Db), and the Symlet wavelet (Sym). In order to yield a thorough and unbiased comparison, we note that the three investigated wavelets were evaluated in the same classification conditions, i.e., the same number of features, the same training beats, the same classifier and the same test beats. In general, as can be observed from Fig. 4a, these results show the superiority of the proposed method over the other two wavelets whatever the fil-

ter order and decomposition level adopted. Indeed, in terms of OA, the worst decomposition level of the proposed method (i.e., level 1) outperforms the best decomposition level for Daubechies and Symlet wavelets (i.e., level 4). As it can be expected, Fig. 4b suggests clearly that classification-driven wavelets (PSO-based wavelets) are significantly more stable (standard deviation=0.81) than the general purpose wavelets for Daubechies and Symlet (standard deviation=2.71 and 2.67), respectively, when dealing with ECG classification problems. In other words, the PSO-based wavelet is less sensitive to the choice of the filter order and decomposition level than standard wavelets are.

The best accuracies achievable by each method are detailed in Table 3. The PSO-based method gives substantial accuracy improvements when compared to standard wavelets. Indeed, the gains in terms of OA (and AA) are 2.18% (1.56%) and 2.01% (2.68%) over the Daubechies and the Symlet wavelets, respectively. The worst class accuracy is obtained on atrial premature beats (A) with accuracies of 82.77%, 84.03%, and 81.09% for PSO, Db, and Sym, respectively. The best class accuracy is yielded on paced beats (/) with accuracies of 96.19%, 92.33%, and 94.55% for PSO, Db, and Sym, respectively. A close inspection of Table 3 shows also that our proposed method does not favor just the dominant class, i.e., normally beats (N), but also at least three other classes among five including classes in the minority. The McNemar's test confirms that the superiority of the proposed method is statistically significant over the other two wavelets (see Table 4).

Table 4

Statistical significance of differences in classification accuracy between the three wavelets expressed by means of the McNemar's test. The differences refer to results reported in Table 3.

	Daubechies	Symlet
PSO-based	19.18	17.63
Daubechies		−1.54

The boldface values represent the McNemar's test value between the proposed method and the standard wavelets.

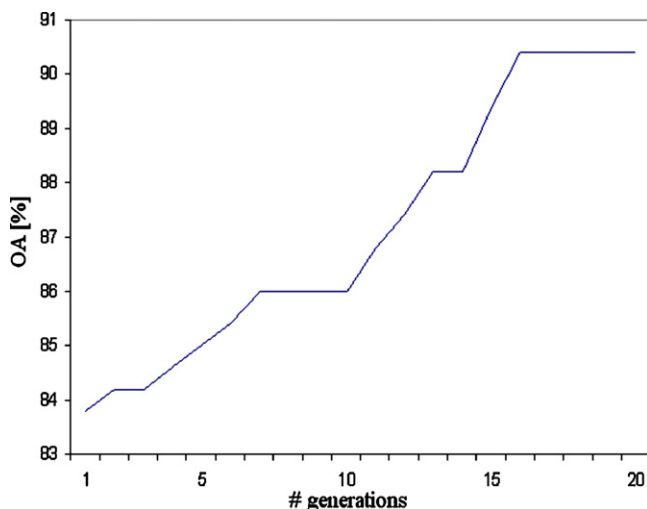


Fig. 3. Example of behavior of the PSO fitness function during the search process.

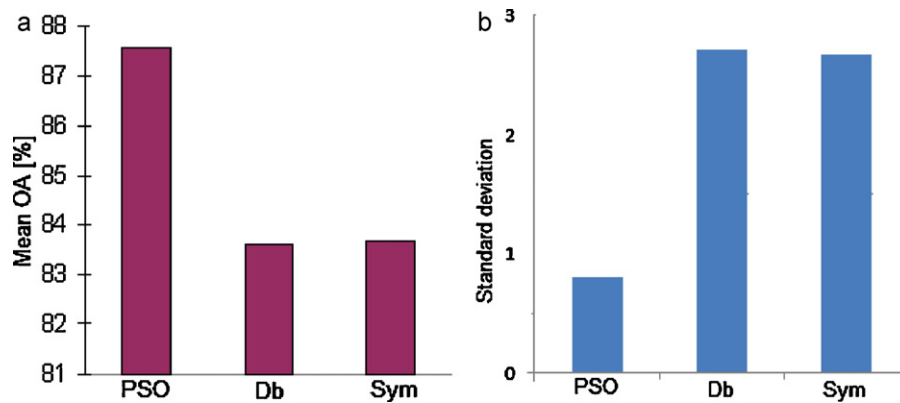


Fig. 4. Mean and standard deviation of the overall accuracy (OA) yielded by the three wavelets by varying the decomposition level (from 1 to 4) and the filter order (from 4 to 10).

Table 5

Sensitivity (*Se*), specificity (*Sp*), and positive predictivity (*Pp*) for ventricular premature beats (V) achieved with the Daubechies, the Symlet, and the PSO-based wavelets.

	<i>Se</i>	<i>Sp</i>	<i>Pp</i>
Daubechies	89.72	94.84	70.54
Symlet	85.40	95.69	71.64
PSO-based	91.75	96.14	74.26

The best classification results are shown in boldface.

Table 5 suggests that the PSO-based wavelet outperforms the standard wavelets in distinguishing ventricular premature beats (V) from the others. In particular, it yielded improvements of 6.35% and 2.03% in sensitivity, 0.45% and 1.30% in specificity, 2.62% and 3.72% in positive predictivity, over the Symlet and Daubechies wavelets, respectively. It is worth noting that these improvements were possible though the wavelet was optimized in a multiclass classification context and not in a dedicated binary classification context (i.e., ‘V’ against all other classes).

7. Conclusion

In this paper, we proposed a novel wavelet optimization procedure based on the combination of the polyphase representation of wavelets and the PSO. It seeks for the wavelet that best represents the beats in terms of discrimination capability measured through an empirical estimate of the classifier accuracy. Therefore, it optimizes the ECG signal representation for the classification task under hand. The procedure was validated on the benchmark MIT/BIH arrhythmia database by adopting as classifier the state-of-the-art SVM classifier. The experimental results show clearly that it achieves better classification accuracies and a higher stability compared to two standard wavelets (Daubechies and Symlet). Though we used the SVM classifier, the procedure is general and thus could be applied as well with any other kind of classification approach. Its main drawback is the substantial processing time required to reach convergence, especially when large training sets are considered.

References

- [1] M. Unser, A. Aldroubi, A review of wavelets in biomedical applications, in: IEEE Proceedings, vol. 84, 1996, pp. 626–638.
- [2] R.S.H. Istepanian, L.J. Hadjileontiadis, S.M. Panas, ECG data compression using wavelets and higher order statistics methods, IEEE Transactions on Information Technology in Biomedicine 5 (2001) 108–115.
- [3] R. Benzid, F. Marir, N. Bouguechal, Electrocardiogram compression method based on the adaptive wavelet coefficients quantization combined to a modified two-role encoder, IEEE Signal Processing Letters (2007) 373–376.
- [4] T. Ince, S. Kiranyaz, M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals, IEEE Transactions on Biomedical Engineering 56 (2009) 1415–1426.
- [5] J.S. Sahambi, S.N. Tandon, R.K.P. Bhatt, Using wavelet transform for ECG characterization, IEEE Engineering in Medicine & Biology 16 (1997) 77–83.
- [6] L.Y. Shyu, Y.H. Wu, W. Hu, Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG, IEEE Transactions on Biomedical Engineering 51 (2004) 1269–1273.
- [7] H. Dickhaus, H. Heinrich, Classifying biosignals with wavelet networks a method for noninvasive diagnosis, IEEE Engineering in Medicine & Biology 15 (1996) 103–111.
- [8] J.S. Lim, Finding features for real time premature ventricular contraction detection using a fuzzy neural network system, IEEE Transactions on Neural Networks 20 (2009) 522–527.
- [9] C. Li, C. Zheng, C. Tai, Detection of ECG characteristics points using wavelet transforms, IEEE Transactions on Biomedical Engineering 42 (1995) 21–28.
- [10] A. Khamene, S. Negahdaripour, A new method for the extraction of fetal ECG from the composite abdominal signal, IEEE Transactions on Biomedical Engineering 47 (2000) 507–516.
- [11] T. Inan, L. Giovannardi, J.T.A. Kovacs, Robust neural network based classification of premature ventricular contractions using wavelet transform and timing interval features, IEEE Transactions on Biomedical Engineering 53 (2006) 2507–2515.
- [12] A.H. Khandoker, M. Palaniswami, C.K. Karmakar, Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings, IEEE Transactions on Information Technology in Biomedicine 13 (2009) 37–48.
- [13] L. Senhaji, G. Carraut, J.J. Bellanger, G. Passariello, Comparing wavelet transforms for recognizing cardiac patterns, IEEE Engineering in Medicine & Biology 14 (1995) 167–173.
- [14] Y. Bazi, F. Melgani, Semisupervised PSO-SVM regression for biophysical parameter estimation, IEEE Transactions on Geoscience and Remote Sensing 45 (2007) 1887–1895.
- [15] M. Donelli, R. Azzaro, F.G.B. De Natale, A. Massa, An innovative computational approach based on a particle swarm strategy for adaptive phased-arrays control, IEEE Transactions on Antennas and Propagation 54 (2006) 888–898.
- [16] F. Melgani, Y. Bazi, Classification of electrocardiogram signals with support vector machines and particle swarm optimization, IEEE Transactions on Information Technology in Biomedicine 12 (5) (2008) 667–677.
- [17] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, Burlington, MA, 1999.
- [18] I. Daubechies, Orthonormal bases of compactly supported wavelets, Communication in Pure Applied Mathematics 41 (1988) 909–966.
- [19] B.G. Sherlock, D.M. Monro, On the space of orthonormal wavelets, IEEE Transactions on Signal Processing 46 (1998) 1716–1720.
- [20] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, 1996.
- [21] P.P. Vaidyanathan, Multirate Systems and Filter Banks, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [22] J. Kennedy, R.C. Eberhart, Swarm Intelligence, Morgan Kaufmann, San Mateo, CA, 2001.
- [23] D.E. Goldberg, Genetic algorithms in search, in: Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989.
- [24] N. Ghogghali, F. Melgani, Y. Bazi, A multiobjective genetic SVM approach for classification problems with limited training samples, IEEE Transactions on Geoscience and Remote Sensing 47 (2009) 1707–1718.
- [25] E. Pasolli, F. Melgani, M. Donelli, Automatic analysis of GPR images: a pattern recognition approach, IEEE Transactions on Geoscience and Remote Sensing 47 (2009) 2206–2217.
- [26] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

- [27] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, *IEEE Transactions on Neural Networks* 13 (2002) 415–425.
- [28] M. Stone, Cross-validatory choice and assessment of statistical predictions, *Journal of Royal Statistics Society B* 36 (1974) 111–147.
- [29] R. Mark, G. Moody, MIT-BIH Arrhythmia Database, 1997 (Online). Available from: <http://ecg.mit.edu/dbinfo.html>.
- [30] F. de Chazal, R.B. Reilly, A patient adapting heart beat classifier using ECG morphology and heartbeat interval features, *IEEE Transactions on Biomedical Engineering* 53 (2006) 2535–2543.
- [31] J.J. Wei, C.J. Chang, N.K. Shou, G.J. Jan, ECG data compression using truncated singular value decomposition, *IEEE Transactions on Biomedical Engineering* 5 (2001) 290–299.
- [32] A. Agresti, *Categorical Data Analysis*, 2nd edition, Wiley, New York, 2002.
- [33] Y.H. Hu, S. Palreddy, W.J. Tompkins, A patient-adaptable ECG beat classifier using a mixture of experts approach, *IEEE Transactions on Biomedical Engineering* 44 (1997) 891–900.