

A New ECG Signal Classification Based on WPD and ApEn Feature Extraction

Hongqiang Li¹ · Xiuli Feng¹ · Lu Cao² · Enbang Li³ ·
Huan Liang¹ · Xuelong Chen¹

Received: 28 June 2014 / Revised: 22 April 2015 / Accepted: 23 April 2015 /
Published online: 5 May 2015
© Springer Science+Business Media New York 2015

Abstract Electrocardiogram (ECG) signal classification is an important diagnosis tool wherein feature extraction plays a crucial function. This paper proposes a novel method for the nonlinear feature extraction of ECG signals by combining wavelet packet decomposition (WPD) and approximate entropy (ApEn). The proposed method first uses WPD to decompose ECG signals into different frequency bands and then calculates the ApEn of each wavelet packet coefficient as a feature vector. A support vector machine (SVM) classifier is used for the classification. The particle swarm optimization algorithm is used to optimize the SVM parameters. The proposed method does not require dimensionality reduction, has fast calculation speed, and requires sim-

✉ Hongqiang Li
lihongqiang@tjpu.edu.cn
Xiuli Feng
1330093069@stu.tjpu.edu.cn
Lu Cao
caolutj@yahoo.com.cn
Enbang Li
enbang_li@uow.edu.au
Huan Liang
1331097101@stu.tjpu.edu.cn
Xuelong Chen
1331097094@stu.tjpu.edu.cn

¹ School of Electronics and Information Engineering, Tianjin Polytechnic University, No. 399 Binshuixi Road, Xiqing District, Tianjin 300387, China

² Tianjin Chest Hospital, Tianjin 300000, China

³ School of Physics, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, NSW 2522, Australia

ple computations. The classification of the signals into five beats yields an acceptable accuracy of 97.78 %.

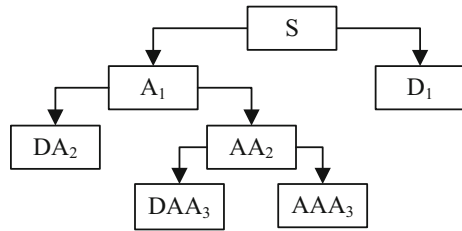
Keywords Approximate entropy · Classification · Feature extraction · Support vector machine · Wavelet packet decomposition

1 Introduction

Electrocardiogram (ECG) signal classification is the most important and common tool for describing and analyzing heart activity. The purpose of feature extraction is to find as few properties of an ECG signal as possible that would facilitate the successful classification of arrhythmia signals and efficient diagnosis. This paper hypothesizes that nonlinear dynamic behavior is a key characteristic of ECG signals and that nonlinear feature extraction is effective for the classification of such. Many methods have been proposed for the feature extraction and classification of ECG signals. A classification was constructed using linear discriminant analysis (LDA) based on the time feature [2]. Zadeh et al. [19] extracted ECG morphological and time interval features and designed several support vector machine (SVM) classifiers that categorized ECG signals into three different classes. Kamath [10] analyzed ECG beats from the perspective of energy by accounting for the features derived from the nonlinear component in the time and frequency domains with the use of the Teager energy operator. Ye et al. [18] constructed a heartbeat classification method that is based on a combination of morphological features extracted by wavelet transform and independent component analysis (ICA) as well as dynamic features derived from RR interval information. A method for fetal ECG feature extraction that uses multidimensional ICA has also been proposed [3]. Fei [6] applied particle swarm optimization (PSO) and SVM to diagnose arrhythmia on the basis of the time domain features. Giri et al. [7] utilized principal component analysis, LDA, ICA, and discrete wavelet transform (DWT) in the automated diagnosis of coronary artery disease. Kutlu and Kuntalp [12] derived features by using the higher order statistics of wavelet packet decomposition (WPD) coefficients and used them as inputs into a classifier based on the k-nearest neighbor algorithm. Nonlinear features, including approximate entropy (ApEn), were extracted from heart rate variability (HRV) and ECG signals with a high classification accuracy [5]. Nonlinear features were used as feature sets to be placed into several classifiers [1].

Unlike in the aforementioned pieces of the literature, the present paper derives nonlinear feature vectors by combining WPD and ApEn. The feature vectors are then used as inputs in the SVM classifier to categorize ECG beats into five types: normal beat, atrial premature beat (APC), supraventricular tachyarrhythmia (SVTA), left bundle branch block beat (LBBB), and premature ventricular contraction (PVC). The rest of this paper is organized as follows. Section 2 presents the proposed method and its theoretical basis. Section 3 discusses the simulation studies conducted to evaluate the proposed method. Section 4 concludes the paper.

Fig. 1 WD structures



2 Proposed Method

The features of each type of ECG signal are derived through the following steps:

Step 1: The preprocessed ECG signals [13, 14] are decomposed into different frequency bands by WPD. The ECG signals are first decomposed by lifting scheme wavelet transform. Then, the wavelet coefficients of the signals are extracted, and an improved half-soft threshold based on the lifting wavelet is used to denoise the ECG signals.

Step 2: The ApEns of the wavelet packet coefficients are calculated. The obtained ApEn is then used as a feature vector and applied as an input into the SVM classifier. The MIT-BIH arrhythmia database is used as the source of ECG records.

2.1 Wavelet Packet Decomposition

WPD is an extension of wavelet decomposition (WD). WPD includes multiple bases, where each base provides a different classification performance and covers the shortage of fixed time–frequency decomposition in DWT [7]. WD splits the original signal into two subspaces, namely A and D, which complement each other. A and D function as spaces that include the low- and high-frequency information of the original signal, respectively. Figure 1 shows the decomposition of the low-frequency subspace. A is repeated, and WD finely partitions the frequency axis toward the low frequency. WPD is a generalized version that decomposes high-frequency bands in WD. WPD provides a more detailed signal decomposition than WD. WPD divides frequency bands into multilevel bands and selects the corresponding band according to the features of the analyzed signal to match the signal spectrum and increase the time–frequency resolution. WPD results in a complete wavelet packet tree (Fig. 2), where $U_{j,n}$ is the n th subspace (n being the frequency factor, $n = 0, 1, 2, \dots, 2j - 1$, and j is the scale factor) of the wavelet packet at the j th scale, and $U_{j,k}^n(t)$ is its corresponding orthonormal basis, where $U_{j,k}^n(t) = 2^{-j/2}u^n(2^{-j}t - k)$ (k is the shift factor). Thus, Eqs. (1) and (2) are satisfied.

When n is even, then

$$u_{j,0}^n(t) = \sum_k h_0(k)u_{j-1,k}^j \tag{1}$$

When n is odd, then

$$u_{j,0}^n(t) = \sum_k h_1(k)u_{j-1,k}^j \tag{2}$$

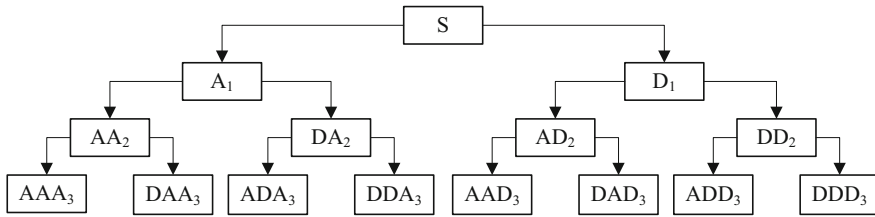


Fig. 2 WPD structures

where $h_0(k)$ and $h_1(k)$ in Eqs. (1) and (2) are several quadruple mirror filters that are irrelevant to scales and satisfies Eq. (3):

$$h_1(k) = (-1)^{1-k}h_0(1 - k) \tag{3}$$

Many different mother wavelets exist, such as Haar, Daubechies (db), Symlet, Coiflet, biorthogonal, and reverse biorthogonal. The wavelet packet coefficients and classification accuracy of the extracted features from each mother wavelet were compared by Kutlu and Kuntalp [12]. They reported that the best feature set can be obtained by using the db6 wavelet function. Thus, the db6 wavelet is selected as the mother wavelet function to estimate the wavelet packet coefficients in the proposed method. The ECG signals are decomposed into three levels. The structures are shown in Fig. 2.

2.2 Approximate Entropy

ApEn is a statistical measure used to quantify data regularities without a priori knowledge of the problem [8]. ApEn adds a real number to a series. A larger real number corresponds to higher series complexity and irregularity. The algorithm that determines the ApEn can be divided into several steps.

1. $N - m$ vectors of dimension m are formed:

$$X(i) = [x(i), x(i + 1), \dots, x(i + m - 1)]$$

$$i = 1, \dots, N - m + 1 \tag{4}$$

These vectors represent m consecutive x signal values that start with i .

2. The distance between $X(i)$ and $X(j)$ is defined as the highest norm:

$$d[X(i), X(j)] = \max_{k=1,2,\dots,m} |x(i + k - 1) - x(j + k - 1)| \tag{5}$$

3. For each $X(i)$, the number of $X(j)$ ($j = 1, \dots, N - m + 1, j \neq i$) is counted such that $d[x(i), x(j)] \leq r$ is satisfied, where r is the tolerance frame parameter. This number is designated as $N^m(i)$. The $B_r^m(i)$ coefficients are then found by using the following expression:

$$B_r^m(i) = \frac{N^m(i)}{N - m + 1} \quad (6)$$

4. The natural logarithms for each $B_r^m(i)$ are calculated, and their mean value is determined as follows:

$$\phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln B_r^m(i) \quad (7)$$

5. The dimension is increased to $m + 1$, and Steps 1–4 are repeated. Thus, $B_r^{m+1}(i)$ and $\phi^{m+1}(r)$ are obtained.
6. The ApEn is found by using the following expression:

$$\text{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (8)$$

Parameters m and r are determined on the basis of a specific problem. m is usually equal to 1 or 2. The range of r varies in series complexity when more points are included. Pincus [16] suggested that $m = 2$ and $r = 0.1 \text{ SD} - 0.2 \text{ SD}$, where SD denotes the standard deviation of the original data $X(i)$. ApEn has a suitable large transient interference capacity because the data for this type of interference (wild points) and adjacent points link into two or three line segments. Moreover, the distance of $X(i)$ is large and must be removed from the threshold value detection. Rather than just describing or reconstructing the strange attractor appearance, ApEn differentiates the complexities of time from a statistical perspective and characterizes the differences or changes in the power system. Thus, ApEn only requires short data (e.g., 1000 points are suitable).

The proposed method calculates the ApEn of each WPD coefficient of 1000 points. Parameters $m = 2$ and $r = 0.2 \text{ SD}$ are adopted in this study.

2.3 Support Vector Machine

The SVM classifier is a supervised learning algorithm that is based on the statistical learning theory developed by Vapnik [17]. This learning algorithm maps low-dimensional datasets to a high-dimensional feature space. The SVM classifier also solves binary problems by searching an optimal hyperplane that can separate the two datasets with the largest margin in the high-dimensional space. The optimal hyperplane is established by a set of support vectors from the original datasets. These subsets form the boundary between two classes. Given the training dataset $\{x_i, y_i\}_{i=1}^N$, where x_i is the input feature vector for each sample and N is the sample number, $y_i \in \{-1, +1\}$ represents its label. Figure 3 shows that the optimal hyperplane is defined as $\omega \cdot x + b = 0$, where x is the point lying on the hyperplane, ω is the parameter for the orientation of the hyperplane, and b is a scalar threshold that represents the bias from both classes. The distance between these two parallel hyperplanes is $1/2\|\omega\|^2$. The input feature vectors satisfy the following inequality:

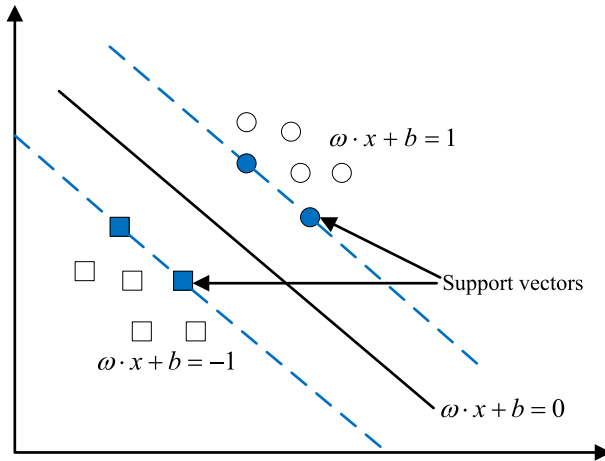


Fig. 3 Data classification using SVM

$$y_i (\omega \cdot x_i + b) \geq 1 \quad i = 1, 2, \dots, N \tag{9}$$

The positive slack variable ξ_i is introduced to obtain the optimal hyperplane and solve the following optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i(\omega \cdot x_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, N, \end{aligned} \tag{10}$$

C is a penalty parameter that controls the trade-off between margin maximization and error minimization. After solving the Lagrange equation of Eq. (10), a classification function can be defined as follows:

$$f(x) = \text{sign} \left\{ \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right\} \tag{11}$$

where α_i is the Lagrange multiplier and $K(x_i, x) = (\varphi(x_i) \cdot \varphi(x))$ is a symmetric and positive kernel function given by Mercer’s theorem. The kernel function can map a low-dimensional vector to a high feature space through nonlinear functions. Three kernel functions are generally used, and their mathematical formulas are provided as follows:

Linear kernel: $K(x_i, x) = (x \cdot x_i)$;

Polynomial kernel: $K(x_i, x) = ((x \cdot x_i) + c)^d, c \geq 0$;

Radial basis function (RBF) kernel: $K(x_i, x) = \exp\left(\frac{-\|x_i - x\|^2}{2\delta^2}\right)$ (δ is a positive real number).

The LIBSVM [4] and popular RBF are adopted in this study. The PSO algorithm is used to determine the optimal parameters of the SVM and RBF function.

Table 1 Training and test sets

Classes	MIT-BIH	Training set	Test set
N	100, 101, 103, 105, 113, 115, 117, 123	30	30
APC	209, 232	30	30
SVTA	207, 209	20	20
LBBB	109, 111	30	30
PVC	200, 201, 223, 228, 233	25	25
	Total	135	135

2.4 Particle Swarm Optimization

PSO performs searches by using a population (or swarm) of individuals (or particles) that are updated from iteration to iteration [11]. Each particle moves in the direction of its previously best position (pbest) and its best global position (gbest) to discover the optimal solution. The velocity and position of particles can be updated by using the following equations:

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 \cdot \text{rand}_1 \cdot (\text{pbest}_{ij}(t) - p_{ij}(t)) + c_2 \cdot \text{rand}_2 \cdot (\text{gbest}_{ij}(t) - p_{ij}(t)) \quad (12)$$

$$p_{ij}(t+1) = p_{ij}(t) + \beta \cdot v_{ij}(t+1) \quad (13)$$

where t denotes the iteration counter, v_{ij} is the velocity of particle i on the j th dimension, whose value is limited to the range $[-v_{\max}, v_{\max}]$; p_{ij} is the position of particle i on the j th dimension, whose value is limited to the range $[-p_{\max}, p_{\max}]$; pbest_{ij} is the pbest position of particle i on the j th dimension; and gbest_{ij} is the gbest position of the swarm on the j th dimension. The inertia weight ω is used to balance the global exploration and local exploitation. rand_1 and rand_2 are random functions in the range $[0, 1]$, whereas β is a constraint factor used to control the velocity weight. Positive constants c_1 and c_2 are personal and social learning factors.

The process of optimizing SVM parameters with PSO is described below:

1. Initialization. PSO is initialized with a population of random particles and velocities. c_1 is set to 1.5, and c_2 is set to 1.7. Max iteration is set to 200, whereas swarm size is set to 20.
2. Training SVM model and fitness evaluation. The SVM model is trained with the parameters C and δ included in the current particle. A global search and iterative optimization are then conducted.
3. Updating pbest_{ij} for each particle and gbest_{ij} for the entire swarm.
4. Updating the velocity and position value of each particle. The velocity of each particle is calculated by using Eq. (12). Each particle moves to its next position according to Eq. (13).
5. Termination. The procedures from Steps 2 to 4 are repeated until the stop conditions are satisfied.

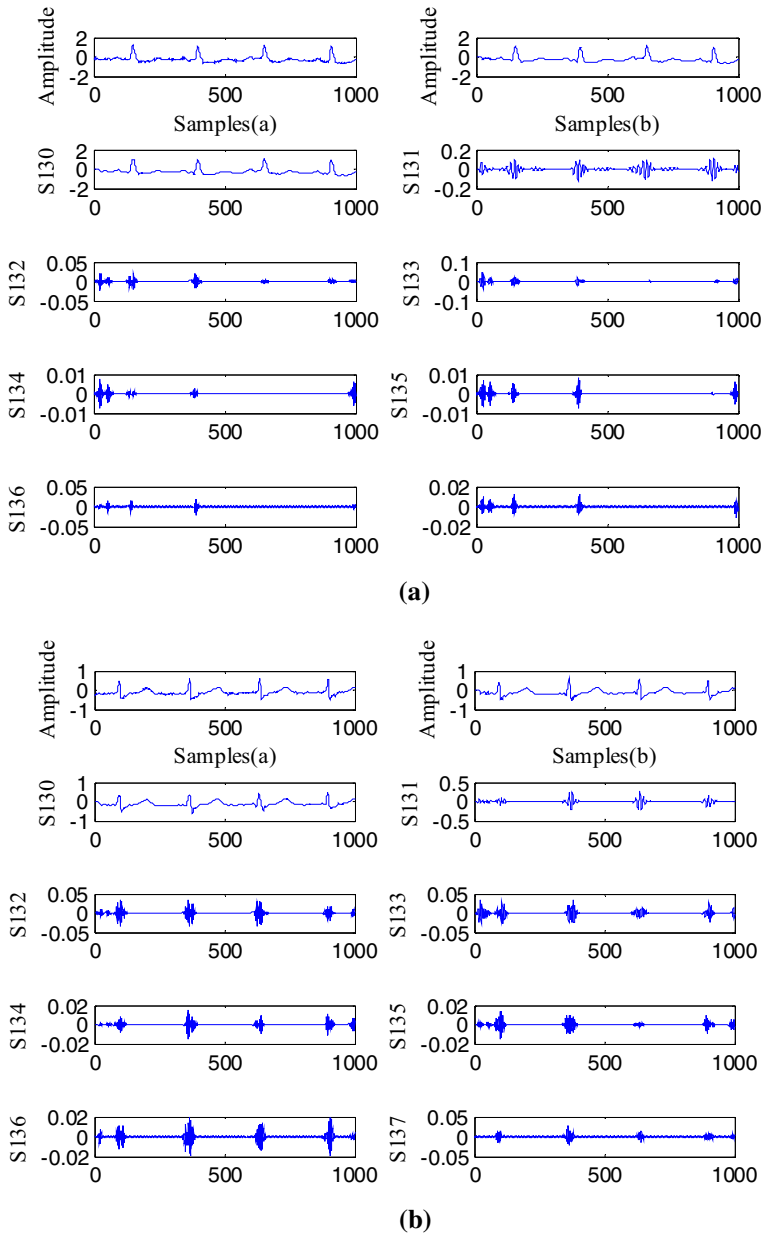


Fig. 4 **a** WPD of the 105 data. **b** WPD of the 232 data. **c** WPD of the 209 data. **d** WPD of the 109 data. **e** WPD of the 228 data

3 Simulation Studies

A simulation study is conducted using the MIT-BIH arrhythmia database to evaluate the performance of the proposed method. The five ECG beats are classified as follows:

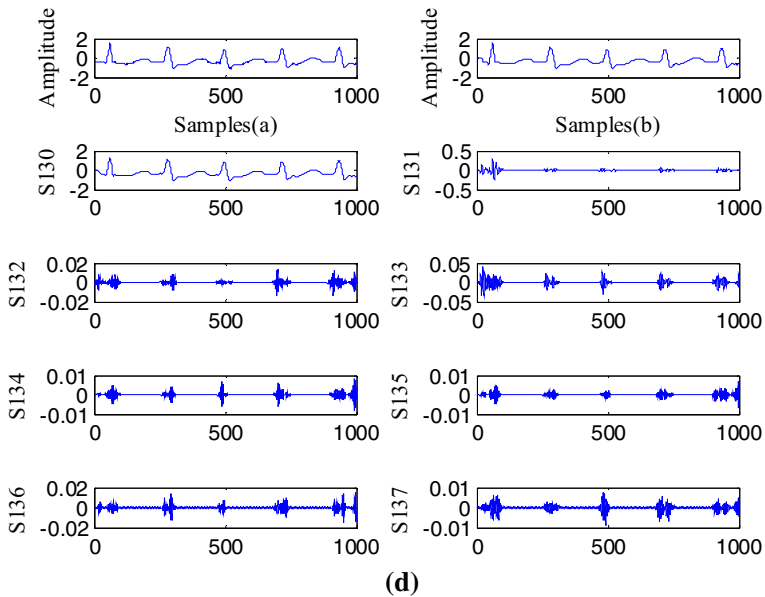
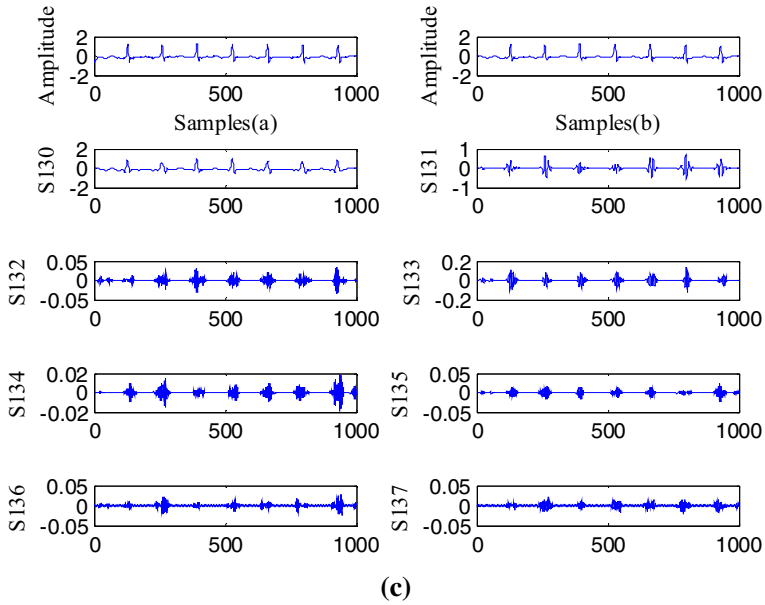


Fig. 4 continued

normal (100, 101, 103, 105, 113, 115, 117, and 123), APC (209 and 232), SVTA (207 and 209), LBBB (109 and 111), and PVC (200, 201, 223, 228, and 233). A total of 270 samples are selected. The training and test sets are shown in Table 1. We select 1000 points of each type that contains the corresponding signal.

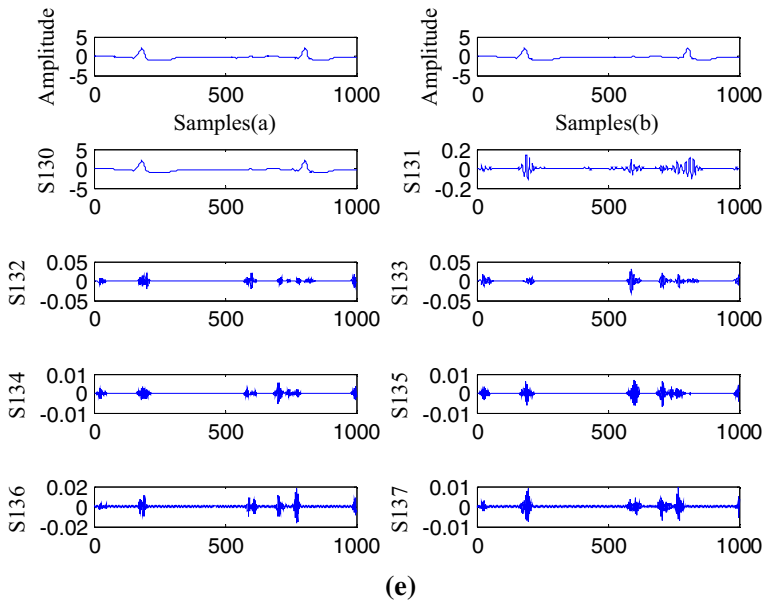


Fig. 4 continued

3.1 Feature Extraction

The nonlinear feature is extracted by WPD and ApEn. The WPD of the ECG signal is shown in Fig. 4, and 105, 232, 209, 109, and 228 data are used as examples on behalf of the five signal types, respectively. Sample (a) is the original signal, whereas sample (b) is the denoised ECG signal. The wavelet packet coefficients, s130 to s137, are used to replace the denoised ECG signal. Figure 4 shows the procedure operation results of the WPD of signals. The wavelet packet coefficients of different signals after decomposition, s130–s137, differ from one another. Thus, the obtained ApEn of the wavelet packet coefficients also have their own characteristics. The method discussed in Sect. 2 is used to calculate the ApEn of every wavelet packet coefficient. The acquired ApEn is considered as a feature. The ApEn of the five classes is listed in Table 2 and used as an example. The mean and standard deviation of each ApEn indicate the significance of that feature's variability in the five beat classes.

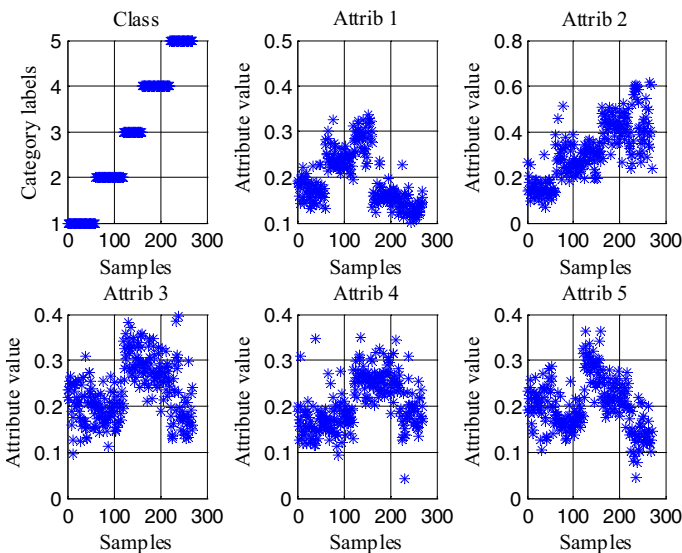
3.2 Classification

The ApEn calculated above is used as an input to the LIBSVM. The labels and distribution of the samples of each category are shown in Fig. 5. Labels 1, 2, 3, 4, and 5 represent N, APC, SVTA, LBBB, and PVC, respectively.

PSO algorithm is used to determine the best parameters. The RBF is generally used because unlike a linear kernel function, it can classify multidimensional data. Additionally, the RBF has fewer parameters to set than a polynomial kernel function.

Table 2 Mean and standard deviation of each ApEn

ApEn	N	APC	SVTA	LBBB	PVC
S130	0.1726 ± 0.0233	0.2416 ± 0.0239	0.2886 ± 0.0269	0.1608 ± 0.0254	0.1367 ± 0.0217
S131	0.1494 ± 0.0402	0.2784 ± 0.0318	0.2950 ± 0.0451	0.4266 ± 0.0554	0.4272 ± 0.1148
S132	0.2076 ± 0.0447	0.1910 ± 0.0309	0.3134 ± 0.0343	0.2749 ± 0.0340	0.2021 ± 0.0632
S133	0.1646 ± 0.0391	0.1725 ± 0.0340	0.2636 ± 0.0293	0.2573 ± 0.0326	0.1856 ± 0.0420
S134	0.2063 ± 0.0423	0.1717 ± 0.0246	0.2889 ± 0.0345	0.2244 ± 0.0280	0.1397 ± 0.0357
S135	0.2091 ± 0.0456	0.1799 ± 0.0293	0.3098 ± 0.0424	0.2520 ± 0.0301	0.1417 ± 0.0350
S136	0.1936 ± 0.0397	0.1531 ± 0.0293	0.2865 ± 0.0352	0.1924 ± 0.0293	0.1214 ± 0.0327
S137	0.2061 ± 0.0396	0.1814 ± 0.0324	0.3224 ± 0.0386	0.2655 ± 0.0348	0.1436 ± 0.0382

**Fig. 5** Category labels and attribute value of samples

Therefore, RBF is a more effective option for kernel function. This study applies an RBF kernel function in the SVM to obtain an optimal solution. Two major RBF parameters (i.e., C and δ) must be set appropriately. The selected value for C influences the classification outcome. A large C results in a high classification accuracy rate in the training phase but a low one in the testing phase. The classification accuracy rate becomes unsatisfactory when C is excessively small, thus rendering the model ineffective. The δ value largely affects the partitioning outcome in the feature space, which affects the classification outcomes. Results are shown in Fig. 6. A total of 20 experiments are conducted, and the results show the following best parameters: $C = 2.0132$ and $\delta = 8.5244$.

The results for the classification are shown in Fig. 7. The samples of the five classes used for classification are 30, 30, 20, 30, and 25. The figure shows that: One beat

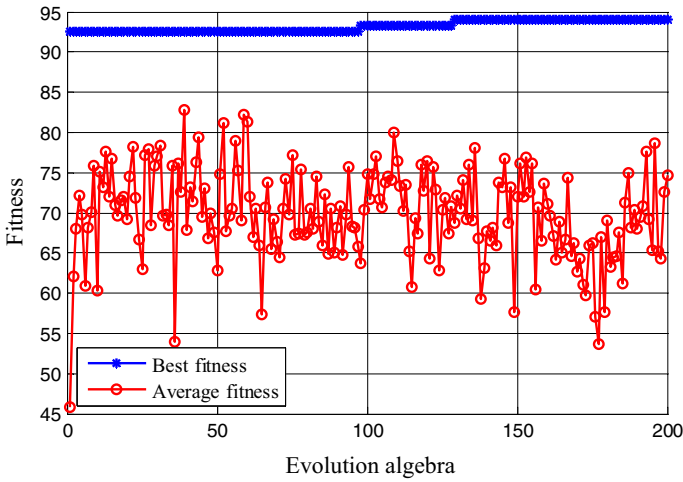


Fig. 6 Fitness curve of the PSO algorithm

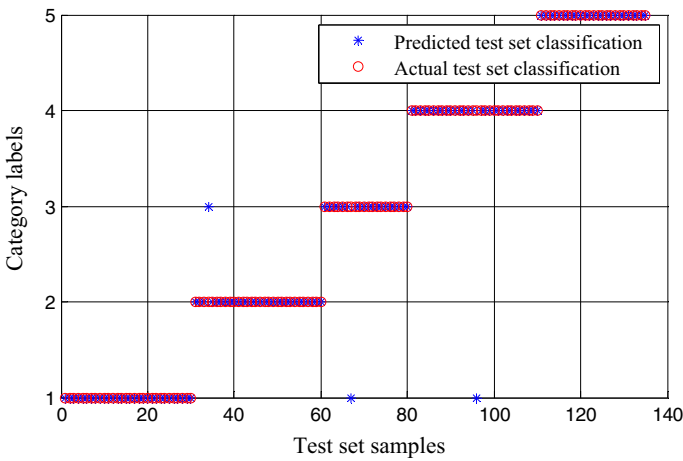


Fig. 7 Actual classification and prediction classification figure based on the PSO algorithm

number of APC is falsely classified into the SVTA, one beat number of SVTA is falsely classified into N, and one beat number of LBBB is falsely classified into N. The results show that the classification accuracy of N, APC, SVTA, LBBB, and PVC is 100, 96.7, 95, 96.7, and 100 %, respectively. The training accuracy is 99.26 %, whereas the total classification accuracy is 97.78 % based on Eq. (14).

$$\text{The classification accuracy} = \frac{\text{The number of correctly classified beats}}{\text{The number of total beats}} \times 100 \% \tag{14}$$

Table 3 Comparison results

Methods	Classifier	Classes	Accuracy (%)	Reference
Morphological and time features	SVM	3	97.14	Zadeh et al. [19]
Teager energy function features	NN	5	95	Kamath [10]
WT and PCA	MD PSO	5	95.58	Ince et al. [9]
Principal components of bispectrum	LSSVM	5	93.48	Martis et al. [15]
Proposed method	LIBSVM	5	97.78	–

3.3 Comparison

A comparison of the proposed method with several other methods is summarized in Table 3. The proposed method achieves a comparable or higher accuracy than the other methods. The proposed method has its own characteristics of no dimensionality reduction and requires only a few samples, thereby reducing the time needed for feature extraction, cross-optimization, and training.

4 Conclusions

This paper presents a novel framework for classifying ECG signals on the basis of nonlinear feature extraction by combining WPD and ApEn. A total of 270 datasets are simulated, and the LIBSVM toolbox is used for classification. PSO algorithm is used to determine the best parameters. The training accuracy is 99.26 %, whereas the classification accuracy is 97.78 %. The simulation results indicate that the proposed method yields an acceptable accuracy without dimensionality reduction and that it can be used as a tool for diagnosis.

Acknowledgments This paper is supported by the National Natural Science Foundation of China (Nos. 61177078, 61307094, 31271871), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20101201120001), and Tianjin Research Program of Application Foundation and Advanced Technology (No. 13JCYBJC16800).

References

1. U.R. Acharya, O. Faust, N.A. Kadri, J.S. Suri, W. Yu, Automated identification of normal and diabetes heart rate signals using nonlinear measures. *Comput. Biol. Med.* **43**(10), 1523–1529 (2013)
2. A.S. Alvarado, C. Lakshminarayan, J.C. Principe, Time-based compression and classification of heartbeats. *IEEE Trans. Biomed. Eng.* **59**(6), 1641–1648 (2012)
3. J.L. Camargo-Olivares, R. Martín-Clemente, S. Hornillo-Mellado, M.M. Elena, I. Román, The maternal abdominal ECG as input to MICA in the fetal ECG extraction problem. *IEEE Signal Process. Lett.* **18**(3), 161–164 (2011)
4. C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines. *ACM TIST* **2**(3), 27 (2011)
5. F. Ebrahimi, S.K. Setarehdan, J. Ayala-Moyeda, H. Nazeran, Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals. *Comput. Methods Prog. Biomed.* **112**(1), 47–57 (2013)
6. S. Fei, Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine. *Expert Syst. Appl.* **37**(10), 6748–6752 (2010)

7. D. Giri, U.R. Acharya, R.J. Martis, S.V. Sree, T.C. Lim, T. Ahamed, J.S. Suri, Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform. *Knowl. Based Syst.* **37**, 274–282 (2013)
8. R. Hornero, D. Abásolo, N. Jimeno, C.I. Sánchez, J. Poza, M. Aboy, Variability, regularity, and complexity of time series generated by schizophrenic patients and control subjects. *IEEE Trans. Biomed. Eng.* **53**(2), 210–218 (2006)
9. T. Ince, S. Kiranyaz, M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals. *IEEE Trans. Biomed. Eng.* **56**(5), 1415–1426 (2009)
10. C. Kamath, ECG beat classification using features extracted from Teager energy functions in time and frequency domains. *IET Signal Process.* **5**(6), 575–581 (2011)
11. J. Kennedy, R. Eberhart, Particle swarm optimization, in *Proceedings of IEEE International Conference on Neural Networks* (1995), pp. 1942–1948
12. Y. Kutlu, D. Kuntalp, Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients. *Comput. Methods Prog. Biomed.* **105**(3), 257–267 (2012)
13. H.Q. Li, X.F. Wang, Detection of electrocardiogram characteristic points using lifting wavelet transform and Hilbert transform. *Trans. Inst. Meas. Control* **35**(5), 574–582 (2013)
14. H.Q. Li, X.F. Wang, L. Chen, E.B. Li, Denoising and R-Peak detection of electrocardiogram signal based on EMD and improved approximate envelope. *Circuits Syst. Signal Process.* **33**(4), 1261–1276 (2014)
15. R.J. Martis, U.R. Acharya, K.M. Mandana, A.K. Ray, C. Chakraborty, Cardiac decision making using higher order spectra. *Biomed. Signal Process.* **8**(2), 193–203 (2013)
16. S.M. Pincus, Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **88**(6), 2297–2301 (1991)
17. V.N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995)
18. C. Ye, B.V.K.V. Kumar, M.T. Coimbra, Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans. Biomed. Eng.* **59**(10), 2930–2941 (2012)
19. A.E. Zadeh, A. Khazaei, V. Ranaee, Classification of the electrocardiogram signals using supervised classifiers and efficient features. *Comput. Methods Prog. Biomed.* **99**(2), 179–194 (2010)