SIViP (2014) 8:111–120
DOI 10.1007/s11760-013-0478-6

ORIGINAL PAPER

A low-complexity data-adaptive approach for premature ventricular contraction recognition

Peng Li · Chengyu Liu · Xinpei Wang · Dingchang Zheng · Yuanyang Li · Changchun Liu

Received: 14 February 2012 / Revised: 26 October 2012 / Accepted: 10 April 2013 / Published online: 26 April 2013

Abstract Premature ventricular contraction (PVC) may lead to life-threatening cardiac conditions. Real-time automated PVC recognition approaches provide clinicians the useful tools for timely diagnosis if dangerous conditions surface in their patients. Based on the morphological differences of the PVC beats in the ventricular depolarization phase (QRS complex) and repolarization phase (mainly T-wave), two beat-to-beat template-matching procedures were implemented to identify them. Both templates were obtained by a probability-based approach and hence were fully data-adaptive. A PVC recognizer was then established by analyzing the correlation coefficients from the two template-matching procedures. Our approach was trained on 22 ECG recordings from the MIT-BIH arrhythmia database (MIT-BIH-AR) and then tested on another 22 nonoverlapping recordings from the same database. The PVC recognition accuracy was 98.2 %, with the sensitivity and positive predictivity of 93.1 and 81.4 %, respectively. To evaluate its robustness against noise, our approach was applied again to the above testing set, but this time, the ECGs were not preprocessed. A comparable performance was still obtained. A good generalization capability was also confirmed by validating our approach on an independent St. Petersburg Institute of Cardiological Technics database. In addition, our performance was comparable with these published complex approaches. In conclusion, we have developed a low-complexity data-adaptive PVC recognition approach with good robustness against noise and generalization capability. Its performance is comparable to other state-of-the-art methods, demonstrating a good potential in real-time application.

Keywords Electrocardiogram (ECG) · Premature ventricular contraction (PVC) · Low-complexity · Data-adaptive · Template matching

P. Li · Chengyu Liu · X. Wang · Changchun Liu (✉)
School of Control Science and Engineering, Shandong University, 17923 Jingshi Road, Jinan 250061, People's Republic of China
e-mail: changchunliu@sdu.edu.cn

P. Li
e-mail: lskyp@mail.sdu.edu.cn

Chengyu Liu
e-mail: bestlcy@sdu.edu.cn

X. Wang
e-mail: wangxinpei@sdu.edu.cn

D. Zheng
Medical Physics, Institute of Cellular Medicine, Freeman Hospital, Newcastle University, Newcastle upon Tyne NE7 7DN, UK
e-mail: dingchang.zheng@newcastle.ac.uk

Y. Li
Department of Medical Engineering, Provincial Hospital Affiliated to Shandong University, 9677 Jingshi Road, Jinan 250014, People's Republic of China
e-mail: sikao666@163.com

1 Introduction

Portable ambulatory ECG monitoring has attracted more and more attentions because it well meets the growing clinical needs for personal healthcare, remote monitoring, point-of-care diagnosis, etc. [1,2]. It demands for fast ECG signal analysis methods so as to help clinicians to diagnose dangerous cardiac conditions in real-time. The immediate detection of life-threatening cardiac arrhythmias is exactly one critical aspect, premature ventricular contraction (PVC) in particular, which has already been showed to be linked to mortality when associated with myocardial infarction [3]. Also, rejecting PVC beats is one essential step for heart rate variability analysis [4,5].

However, most recent studies focused on off-line algorithms, and thus, they employed many complicated mathematical tools (wavelet transform, nonlinear complexity measures, complicated artificial neural network, etc.) [6–16]. Their efficiency is generally accompanied by long computational time and high complexity. While portable ambulatory ECG device is usually battery-driven and runs under an embedded development environment, their quite limited computation resources require low-complexity real-time procedures.

There are also a couple of studies which reported real-time approaches [17,18]. But their high accuracies were based on relatively small data sets (Zhou [17] tested on 20 patients and Hu [18] on 16 recordings). Their efficiency over a large number of files is generally a difficult problem to address. Therefore, the generalization capability of the signal processing algorithms should also be seriously considered [6,12,16]. One approach is to increase the amount of training data. However, no matter how large the data set is, it is still not possible to cover all the potential ECG features of the patients, and the complexity of the classifier grows as the size of the training data set grows [6]. A reasonable generalization capability has been achieved by using a patient-adaptable procedure, which combined a conventional global classifier with a customized classifier developed from patient-specific ECG data or on experts' validation [6,12]. However, this approach is again accompanied with high computational complexity.

The aim of this study is thus to develop a low-complexity data-adaptive PVC recognition approach. Its robustness against noise and generalization capacity will be assessed, and its performance will also be compared with some published off-line high-complexity approaches.

2 Methods

This study included four main stages, which are briefly summarized in Fig. 1. The first stage was the training and testing processes. The proposed algorithm was trained and tested on two un-overlapping data sets (DS1 and DS2) from the MIT-BIH arrhythmia database (MIT-BIH-AR) [19,20]. The second stage was to assess the robustness of algorithm against noise. The PVC recognition was applied to the un-preprocessed testing data set (DS2), with its performance compared with that from the testing process in the first stage. Next, to determine the generalization capability of the algorithm, the developed algorithm was tested on another independent database (the St. Petersburg Institute of Cardiological Technics (INCART) [19]). Finally, the algorithm performance was compared with some previously published studies [10,13,14,16].

2.1 ECG databases

The well-known MIT-BIH-AR database was used for training and testing purposes, and the INCART database for validation. Note that only recordings of lead MLII or lead II in the two databases were used in this study. All the databases are freely available on PhysioNet [19], and their details are briefly described below.

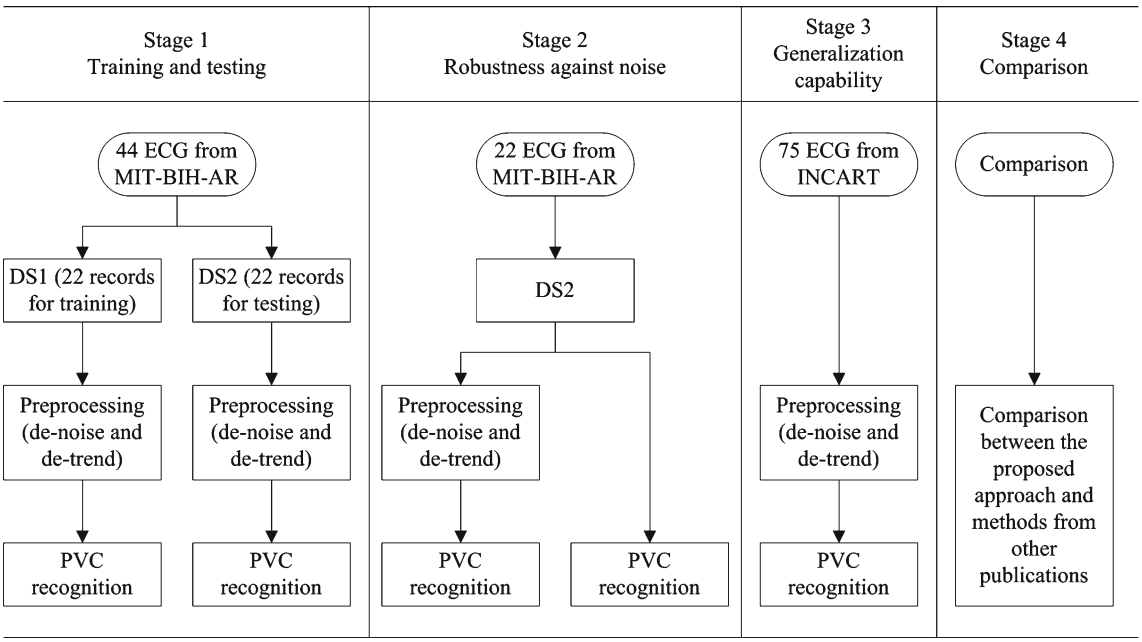

Fig. 1 Schematic representation of the different stages of this study

2.1.1 MIT-BIH arrhythmia database (MIT-BIH-AR)

This database is comprised of 48 half-hour ECG recordings sampled at 360 Hz. Four recordings (102, 104, 107 and 217) with paced beats were discarded. The remaining 44 recordings were divided into two sets—a training set (DS1) and a testing set (DS2). For comparison purpose, DS1 and DS2 were the same as in the published studies [9,12,16]. The ventricular escape beats were also classified as the PVC beats in this study, and the ventricular flutter wave and the unclassifiable beats were discarded [9,16]. The classification scheme and the number of PVC and non-PVC beats in DS1 and DS2 are summarized in Table 1.

2.1.2 INCART 12-lead arrhythmia database

This database consists of 75 annotated recordings extracted from 32 Holter records. They are sampled at 257 Hz. Each recording is 30 min long and contains 12 standard leads. There are over 175,000 beat annotations, which were produced by an automatic algorithm and then corrected manually [19]. None of the recordings has paced beats, and most of them have PVC beats. The annotated number of PVC and non-PVC beats from this database is also given in Table 1.

2.2 Approaches to recognize PVC

2.2.1 Signal preprocessing

The ECG recordings of the INCART database were resampled to 360 Hz with a tenth-order low-pass finite-impulse response (FIR) filter [16] to match the MIT-BIH-AR database, hence to simplify the subsequent preprocessing procedures. All the recordings from both the MIT-BIH-AR and the resampled INCART databases were then preprocessed using median filter to remove the baseline wander and then a 12-tap low-pass FIR filter to remove the unwanted power line and high-frequency noise [9].

2.2.2 Data-adaptive procedure to determine two matching templates

Template matching has been widely used to distinguish the isolated waveforms from a cluster of similar ones. The key to differentiating PVC beats from non-PVC beats lies in the morphological differences in both the ventricular depolarization phase (QRS complex) and repolarization phase (mainly the T-wave). For the PVC beats, the amplitude of the QRS complex is either much larger or lower than that of a normal sinus beat, and there is obvious distortion in the T-wave. Based on these differences, a normal QRS template (T1) and a normal cycle template (T2) were used to identify PVC beats. The two templates were determined by a probability-based data-adaptive approach from the first 5 min of each ECG recording (learning period of the algorithm). The detailed procedure is described below.

(1) Determine the normal beat-to-beat interval range
The beat-to-beat time interval series was constructed from the time difference between two consecutive annotated fiducial points. Extremities were removed by the 3σ criterion. The time interval series was then sorted in ascending order and divided into 5 successive subsets with equal time duration. The number of points in each subset was counted, and the normal beat-to-beat interval range was determined from the subset having the maximum points.

Table 1 Classification scheme and the number of PVC and non-PVC beats in the MIT-BIH-AR and the INCART databases

Dataset	Recordings	Non-PVC beats	PVC beats	Total
DS1	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230	47070	3782	50852
DS2	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234	46335	3213	49548
INCART	I01–I75	155881	20011	175892
	–	Non-PVC	PVC	Discarded
Annotation[a]	–	$N, L, R, a, F, J, A, S, j, e, n$[b]	V, E[c]	Others

[a] Annotated by expert cardiologist. Paced beat, ventricular flutter wave and unclassifiable beats are discarded
[b] $N, L, R, a, F, J, A, S, j, e$ and n refer to *normal beat*, *left bundle branch block beat*, *right bundle branch block beat*, *aberrated atrial premature beat*, *fusion of ventricular and normal beat*, *nodal premature beat*, *atrial premature beat*, *supraventricular premature beat*, *nodal escape beat*, *atrial escape beat and supraventricular escape beat*, respectively
[c] V and E refer to *PVC* and *ventricular escape beat*, respectively

(2) Determine the normal waveform amplitude range
The determination of normal waveform amplitude range was the same as that for the normal beat-to-beat interval range. The amplitude (the absolute amplitude of the fiducial point) series was sorted in ascending order and then divided into 5 successive subsets with equal amplitude range. The normal waveform amplitude range was finally determined from the subset having the maximum points.

(3) Construct T1 and T2
The first beat that fell within the above two normal ranges was identified with its corresponding start and end points of the QRS determined using a derivative-based method. The ECG episode between the start and end points of the QRS was regarded as T1 for that recording, and T2 was determined from the episode between the two fiducial points from the above first beat and its following beat.

2.2.3 PVC recognition by template matching

Beat-to-beat template matching was operated every 5 s ECG with an overlap of 1 s using a normalized correlation coefficient as Eq. (1).

$$x_k \text{ (or } y_k) = \frac{\sum_{n=0}^{L-1} \left[b_k(n) - \overline{b}_k\right]\left[N(n) - \overline{N}\right]}{\sqrt{\sum_{n=0}^{L-1} \left[b_k(n) - \overline{b}_k\right]^2 \left[N(n) - \overline{N}\right]^2}} \quad (1)$$

(1) For x, $b_k(n)$ is the QRS complex in the kth beat, L is the length of the predetermined T1 ($N(n)$ in the equation), $\overline{b}_k$ and $\overline{N}$ are the average of $b_k(n)$ and $N(n)$, respectively. Note that when performing this matching, the fiducial point of each beat was aligned to that of T1, and the start and end of the QRS complex of that beat was determined according to the length of T1. Therefore, the length of $b_k(n)$ is also L.

(2) For y, $b_k(n)$ is the episode between the fiducial points of the kth and $k+1$th beats, L is the length of T2 ($N(n)$ in the equation), and $\overline{b}_k$ and $\overline{N}$ are the average of $b_k(n)$ and $N(n)$, respectively. However, the length of $b_k(n)$ is not always the same to L (due to heart rate variability), and an interval stretching-compressing process [21,22] was thus applied. Let L_k be the length of $b_k(n)$ in the kth beat. The ratio of L_k and L is defined as the compressing-stretching factor α_k. For each beat, a compressing-stretching waveform $\tilde{b}_k(n)$ is calculated as:

$$\tilde{b}_k(n) = b_k[1 + \alpha_k(n-1)] \quad (2)$$

wherein $n = 1, 2, \ldots, L$, and the value for $b_k[1+\alpha_k(n-1)]$ is determined as the interpolation between $b_k[\lfloor 1+\alpha_k(n-1)\rfloor]$ and $b_k[\lceil 1+\alpha_k(n-1)\rceil]$ or the extrapolation if $1+\alpha_k(n-1)$ is larger than L_k. $\tilde{b}_k(n)$ is used as a substitution of $b_k(n)$ in Eq. (1).

The correlation coefficient between the PVC beats and T1 (or T2) is low. Let the z-axis in a Cartesian coordinate system indicates the quantized class index which has a higher value if the beat is more similar to a non-PVC beat and a lower one if it is more similar to a PVC beat. The outcomes of the two template-matching procedures act as the x-axis and y-axis. The projections on both z versus x plane and z versus y plane are monotonously increased. Moreover, we preferred to range the beat into PVC class as long as one coefficient (matching with T1 or T2) is lower. Thus, the slope of the projections should be increased with the increase in x or y. The exponential function is a good representation for both projections and the classification function $z = f(x, y)$, which could be thus expressed as:

$$z_k = f(x_k, y_k) = \frac{e^{x_k^r} + e^{y_k^r}}{2e} \quad (3)$$

The unknown parameter r determines the increasing rate of the slope. The denominator $2e$ is for normalization. The recognition was then simply fulfilled by a comparison between z_k and a threshold value, denoted as z_{thre}. The kth beat is recognized as a non-PVC beat if $z_k \geq z_{\text{thre}}$, and a PVC beat otherwise. Figure 2 illustrates how the recognizer performs. The ECG episodes are selected from the recording 119 in MIT-BIH-AR.

2.2.4 Determination of the threshold value z_{thre}

The selection of the threshold value z_{thre} plays an important role in developing the above PVC recognizer, which was the main purpose of the training process. The PVC recognition performances in the training data set (DS1) of MIT-BIH-AR were obtained for different r (1, 2, 3, 4, 5 and 10 were assigned in this study) in the classification function and different thresholds. They were then compared with the cardiologist's annotations, with the sensitivity (proportion of true negatives in all non-PVC beats, see Table 2) and specificity (proportion of true positives in all PVC beats) calculated. A receiver operating characteristic (ROC) curve, a graphical plot of the sensitivity and 1-specificity, was then employed to determine the optimal threshold. This procedure was carried out using the SPSS software (version 19, SPSS Inc., USA).

Table 3 shows the recognition performance in DS1 with different values of r. The optimal threshold value for a certain r value is also given. Overall, the most acceptable performance was obtained with $r = 4$ or 5. Their corresponding optimal thresholds z_{thre} were 0.59 and 0.55, respectively. The optimal threshold z_{thre} of 0.55 was finally chosen in this study, which is also shown in Fig. 2.

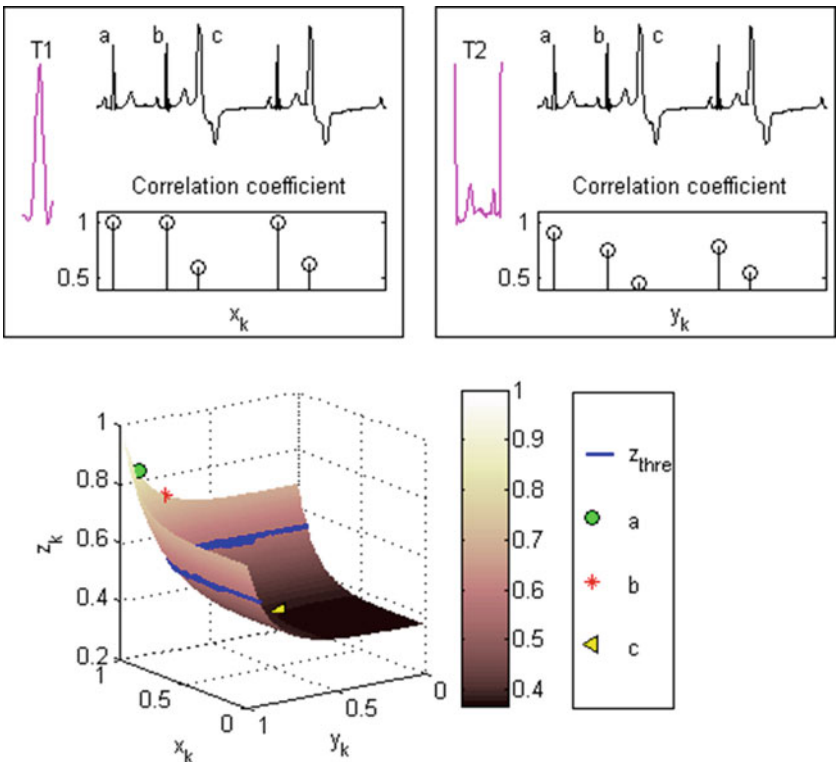

Fig. 2 Illustration of the PVC recognizer. The lower panel shows the classification function as in Eq. (3) when $r = 5$. The ECG episode was from recording 119 in MIT-BIH-AR. Three beats (a, b, c) in the ECG episode and their projections in the classification function are marked

Table 2 Performance measures used in this study

	Our approach		Measures		
	n	v	Total	Non-PVC beats	PVC beats
Annotation					
N	Nn (TN)	Nv (FP)	$A = \frac{Nn + Vv}{Nn + Nv + Vn + Vv}$	$\mathrm{Se} = \frac{Nn}{Nn + Nv}$	$\mathrm{Se} = \frac{Vv}{Vn + Vv}$
V	Vn (FN)	Vv (TP)		$+P = \frac{Nn}{Nn + Vn}$	$+P = \frac{Vv}{Nv + Vv}$

N Non-PVC beats (true), V PVC beats (true), n non-PVC beats (recognized), v PVC beats (recognized), *Nn (TN)* non-PVC beats recognized as non-PVC beats (true negatives), *Nv (FP)* non-PVC beats recognized as PVC beats (false positives), *Vn (FN)* PVC beats recognized as non-PVC beats (false negatives), *Vv (TP)* PVC beats recognized as PVC beats (true positives), A Accuracy, *Se* sensitivity, $+P$ positive predictivity

2.3 Performance quantification

The performance of the proposed recognizer was quantified by accuracy (A), sensitivity (Se) and positive predictivity ($+P$). Table 2 shows how they are calculated in this study.

2.4 Validation and comparison

2.4.1 Validation of the proposed recognizer

After the training process, as shown in Fig. 1, the performance of the developed recognizer was evaluated in the DS2 of the MIT-BIH-AR and the INCART database. The independent INCART database was used to assess the generalization capability of the PVC recognizer. Additionally, DS2 was also reused to test the ability of robustness against the noise of the PVC recognizer.

2.4.2 Comparison with published studies

Our performance results were finally compared with 4 published studies where the entire or part of MIT-BIH-AR and INCART databases were used [10,13,14,16]. The ECG recordings in common with [10,13,14,16] were used, respectively. Their performance results from our approach were compared directly with the corresponding reported results.

Table 3 Overall performance in the training set (DS1) using different values of r in the classification function (z_{thre} provided here is the optimal threshold value at a certain r)

r	z_{thre}	A	Non-PVC beats		PVC beats	
			Se	$+P$	Se	$+P$
1	0.82	92.1	92.0	99.5	93.7	48.4
2	0.72	92.3	92.0	99.6	95.6	49.0
3	0.64	93.2	93.0	99.6	95.0	52.3
4	0.59	93.2	93.0	99.6	95.0	52.4
5	0.55	93.2	93.1	99.6	95.2	52.5
10	0.45	93.0	92.8	99.5	94.7	51.5

The performances are expressed in percentages
A Accuracy, Se sensitivity, $+P$ positive predictivity

3 Results

Table 4 shows the PVC recognition performance for each recording from DS1 and DS2. Table 5 shows the overall validation results of the developed approach, including the results from the testing process using DS2 [Table 5(a)], the assessment of the robustness against noise using un-preprocessed DS2 [Table 5(b)] and the generalization evaluation using the INCART database [Table 5(c)]. An overall accuracy (A) was 98.2 %, with sensitivity (Se) 98.5 % and positive predictivity ($+P$) 99.5 % for non-PVC beats, and Se 93.1 % and $+P$ 81.4 % for PVC beats in DS2. For un-preprocessed DS2, a small nonsignificant decrease was showed (A of 98.0 %; Se 98.4 % and $+P$ 99.4 % for non-PVC beats, and Se 92.0 % and $+P$ 79.4 % for PVC beats). And comparable performances were obtained from INCART database (A of 94.0 %; Se 94.0 % and $+P$ 99.1 % for non-PVC beats, and Se 93.4 % and $+P$ 66.5 % for PVC beats).

Table 6 shows the comparison with 4 published studies [10,13,14,16], where the high-complexity approaches were used. Again, with a much simpler approach in this study, the comparable results were also obtained.

4 Discussion and conclusion

A low-complexity data-adaptive approach for PVC recognition has been proposed in this study. Computerized algorithms for PVC recognition have been well established for a few decades. To improve the accuracy, many complicated mathematical tools have been investigated, and currently, they are almost the dominant approaches. The computational complexity might not be the key issue for off-line ECG analysis. However, with the increasing demand of the remote healthcare monitoring, the low complexity is required in order to make real-time ECG acquisition and analysis feasible. The low-complexity approach proposed here well meets the clinical needs. In addition, generalization capability should also be seriously considered. We believed that it could be improved with a more adaptive procedure. The results of this study validated this hypothesis.

Two simple beat-by-beat template-matching processes were employed to consider the ECG waveform morphological differences between PVC and non-PVC beats. The first one checks the QRS similarity and the second one mainly the T-wave similarity. In theory, the correlation coefficients from the two template-matching processes are very low if it is a PVC beat and vice versa. The classification function expressed in Eq. (3) shows a mutual mapping relationship between the templates and the PVC and non-PVC beats. An exponential map for the classification was used, and a parameter r as given in Eq. (3) was used to adjust the slope between the PVC region and the non-PVC region. Different values of r were assessed in the training process. With $r \geq 4$, their performance varied slightly, and all of them were indeed acceptable. In this study, $r = 5$ was finally selected.

It has also been observed that the PVC recognizer performed better in the testing data set (DS2) than that in the training set (DS1) in terms of A, Se for non-PVC beats and $+P$ for PVC beats, while it was completely comparable in terms of $+P$ for non-PVC beats and Se for PVC beats. This phenomenon agreed with what has been reported in [9,16]. It could be caused by a couple of reasons. Firstly, different noise distribution and different amount of left bundle branch block beats and right bundle branch block beats from the two data sets might account for this bias. Secondly, the majority of PVC beats were distributed in a few recordings in DS2. However, it was much more spread in DS1.

The proposed approach has also been validated on an independent INCART database. The results from the INCART database suggested that our approach have a good generalization capability. It could owe much to the fully data-adaptive procedures we employed. The performance was comparable to that in DS1, but a significant increase in $+P$ for PVC beats was observed. This could be explained by the increased class imbalance of PVC beats in the DS1, which is about 12-to-1, while in INCART database, it is approximately 7-to-1.

Table 4 Performance for each individual recording from DS1 and DS2

Rec.	# Beats		Measures[a]				
	N	*V*	*A*	Non-PVC beats		PVC beats	
				Se	$+P$	Se	$+P$
DS1							
101	1858	0	100	100	100	–	–
106	1503	518	98.8	98.9	99.9	98.7	96.8
108	1740	17	67.8	67.6	99.8	82.4	2.4
109	2487	37	99.9	100	99.9	91.9	100
112	2531	0	100	100	100	–	–
114	1829	43	99.7	99.8	100	97.7	91.3
115	1946	0	100	100	100	–	–
116	2295	109	99.0	99.0	100	99.1	82.4
118	2255	16	100	100	100	100	94.1
119	1537	444	100	100	100	100	100
122	2468	0	100	100	100	–	–
124	1566	47	99.9	99.8	100	100	94.0
201	1761	198	96.2	95.7	100	100	92.5
203	2525	444	71.1	66.4	99.5	98.2	34.0
205	2577	71	100	100	100	100	98.6
207	1640	210	13.5	5.2	65.2	78.1	9.6
208	1955	989	95.2	92.9	99.9	99.8	87.7
209	2996	1	99.9	99.9	100	100	33.3
215	3189	164	99.3	99.8	99.5	89.6	96.1
220	2041	0	100	100	100	–	–
223	2124	473	95.6	99.1	95.7	79.9	95.0
230	2247	1	100	100	100	100	100
Total	47070	3782	93.2	93.1	99.6	95.2	52.5
DS2							
100	2264	1	100	100	100	100	100
103	2078	0	100	100	100	–	–
105	2519	41	94.4	95.6	98.7	22.0	7.4
111	2117	1	87.4	87.4	100	100	0.4
113	1789	0	100	100	100	–	–
117	1530	0	99.9	99.9	100	–	0
121	1855	1	99.7	99.7	100	100	14.3
123	1510	3	100	100	100	100	100
200	1767	826	96.8	99.3	96.2	91.5	98.3
202	2109	19	99.5	99.7	99.9	84.2	69.6
210	2449	193	98.6	99.3	99.9	90.2	90.6
212	2740	0	100	100	100	–	0
213	3021	220	97.9	97.8	99.9	98.6	76.7
214	1997	255	96.8	100	96.5	71.8	99.5
219	2084	63	99.3	100	99.3	77.8	98.0
221	2024	396	99.8	100	99.8	99.0	100
222	2474	0	96.7	96.7	100	–	0
228	1687	360	96.0	95.2	100	100	81.6
231	1563	2	99.1	99.1	100	100	12.5

Table 4 continued

Rec.	# Beats		Measures[a]				
	N	*V*	*A*	Non-PVC beats		PVC beats	
				Se	$+P$	Se	$+P$
232	1776	0	99.8	99.8	100	–	0
233	2240	829	99.6	99.6	99.8	99.5	98.9
234	2742	3	100	100	100	100	100
Total	46335	3213	98.2	98.5	99.5	93.1	81.4

N Non-PVC beats (true), *V* PVC beats (true), *A* Accuracy, *Se* sensitivity, $+P$ positive predictivity, *Rec.* recordings
[a] The measures are expressed in percentages

Table 5 Overall performance in the testing data sets of DS2 and INCART (The performances are expressed in percentages)

		n	*v*	*A*	Non-PVC		PVC	
					Se	$+P$	*Se*	$+P$
(a) DS2	*N*	45653	682	98.2	98.5	99.5	93.1	81.4
	V	221	2992					
(b) DS2 (un-preprocessed)	*N*	45570	765	98.0	98.4	99.4	92.0	79.4
	V	261	2952					
(c) INCART	*N*	146467	9414	94.0	94.0	99.1	93.4	66.5
	V	1323	18688					

N Non-PVC beats (true), *V* PVC beats (true), *n* non-PVC beats (recognized), *v* PVC beats (recognized), *A* accuracy, *Se* sensitivity, $+P$ positive predictivity

Table 6 Comparison between this study and [10,13,14,16]

	Recordings	Measures[a]				
		A	Non-PVC		PVC	
			Se	$+P$	Se	$+P$
Shyu [10]	111, 115, 116, 119, 221, 230, 231	97.0	–	–	99.0	–
Our		98.0	98.0	100	99.7	66.0
Inan [13]	100, 101, 103, 105, 106, 109, 112, 113, 114, 115, 116, 118, 119, 121, 122, 123, 200, 201, 202, 203, 205, 208, 210, 212, 213, 214, 215, 219, 220, 221, 223, 228, 230, 231, 233, 234	–	–	–	76.5	85.8
Our		98.2	98.2	99.6	92.5	80.0
Lim [14]	115, 116, 119, 221,230, 231	99.8	–	–	99.2	–
Our		99.7	99.7	100	99.6	79.0
Llamedo [16]	All recordings of DS2	–	–	–	75.0	96.0
Our		98.2	98.5	99.5	93.1	81.4
Llamedo [16]	All recordings from INCART	–	–	–	82.0	88.0
Our		94.0	94.0	99.1	93.4	66.5

A Accuracy, *Se* sensitivity, $+P$ positive predictivity
[a] The measures are expressed in percentages

The performance obtained in un-preprocessed DS2 data set was slightly decreased, but it was still highly comparable with that after a preprocessing procedure. It suggests that our approach is fairly robust against noise. However, template matching itself is prone to be affected by waveform distortion, so application in recordings with very low

signal-to-noise ratio might not result in good performance. For instance, the recordings 105, 108 and 121 have a very low SNR, and the noise highly distorts the waveforms of non-PVC beats. Consequently, low template-matching results for non-PVC beats were obtained, leading to a very low $+P$ for PVC beats (see Table 4).

In comparison with these published approaches [10, 13,14,16] (Table 6), our PVC recognizer showed some degree of improvement, especially in Se for PVC beats. However, one limitation of this study is that $+P$ for PVC beats was not as good as other approaches. High sensitivity of template matching to large waveform distortions might be one dominant reason. However, our additional analysis showed that the two template-matching approaches could improve $+P$ when compared with that from one template-matching procedure alone (only T1, indicating the QRS similarity) (81.4 % against 69.0 % for DS2) without increasing the computational complexity. On the premise of low computational complexity, there might still be room for improvement. The timing features like RR interval used in many published studies [6,9,12,13,16] might help. Also, a more robust classification function needs to be established if there is an increase in the amount of inputs.

It is worth noting that we used the annotated fiducial points in this study. But in order to fully automate the recognition approach, an automated R-peak detection module is required. Recently, great emphasis has already been placed on R-peak detection, and a number of simple schemes have been reported [23–25]. Its error rate is much lower than that of PVC detection. Thus, it is strongly believed that automating the heartbeat detection process would not necessarily degrade the PVC recognition performance.

In conclusion, we have developed a low-complexity PVC recognizer with good robustness against noise and generalization capability. Its performance is highly comparable to other state-of-the-art methods, showing good potential in real-time application.

Acknowledgments We would like to thank Mr. Jian Li and Miss. Xiuhua Ruan of Shandong University for their aids in the ECG data analysis and statistics and Dr. M. Llamedo of the University of Zaragoza for providing the useful MATLAB manuscript for manipulating recordings in the INCART database. We also thank the anonymous reviewers who gave valuable suggestion that has helped us to improve the quality of the manuscript. This work is supported by the Graduate Independent Innovation Foundation of Shandong University (GIIFSDU, yzc12082), the National Natural Science Foundation of China (61201049), the Excellent Young Scientist Awarded Foundation of Shandong Province (BS2012DX019) and the Independent Innovation Foundation of Shandong University (IIFSDU, 2011GN069).

References

1. Clifford, G.D., Moody, G.B.: Signal quality in cardiorespiratory monitoring. Physiol. Meas. **33**(9), E01 (2012)
2. Nikita, K.S., Lin, J.C., Fotiadis, D.I., Arredondo, M.T.: Editorial: special issue on mobile and wireless technologies for healthcare delivery. IEEE Trans. Biomed. Eng. **59**(11), 3083–3089 (2012)
3. Iwasa, A., Hwa, M., Hassankhani, A., Liu, T., Narayan, S.M.: Abnormal heart rate turbulence predicts the initiation of ventricular arrhythmias. Pacing Clin. Electrophysiol. **28**(11), 1189–1197 (2005)
4. Liu, C.Y., Liu, C.C., Shao, P., Li, L.P., Sun, X., Wang, X.P., Liu, F.: Comparison of different threshold values r for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups. Physiol. Meas. **32**(2), 167–180 (2011)
5. Liu, C.Y., Li, L.P., Zhao, L.N., Zheng, D.C., Li, P., Liu, C.C.: A combination method of improved impulse rejection filter and template matching for identification of anomalous intervals in electrocardiographic RR sequences. J. Med. Biol. Eng. **32**(4), 245–250 (2012)
6. Hu, Y.H., Palreddy, S., Tompkins, W.J.: A patient-adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans. Biomed. Eng. **44**(9), 891–900 (1997)
7. Zhang, X., Zhu, Y., Thakor, N.V., Wang, Z.: Detecting ventricular tachycardia and fibrillation by complexity measure. IEEE Trans. Biomed. Eng. **46**(5), 548–555 (1999)
8. Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., Sornmo, L.: Clustering ECG complexes using Hermite functions and self-organizing maps. IEEE Trans. Biomed. Eng. **47**(7), 838–848 (2000)
9. de Chazal, P., O'Dwyer, M., Reilly, R.B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans. Biomed. Eng. **51**(7), 1196–1206 (2004)
10. Shyu, L.Y., Wu, Y.H., Hu, W.: Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG. IEEE Trans. Biomed. Eng. **51**(7), 1269–1273 (2004)
11. Sun, Y., Chan, K.L., Krishnan, S.M.: Life-threatening ventricular arrhythmia recognition by nonlinear descriptor. Biomed. Eng. Online **4**(1), 6 (2005)
12. de Chazal, P., Reilly, R.B.: A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. IEEE Trans. Biomed. Eng. **53**(12), 2535–2543 (2006)
13. Inan, O.T., Giovangrandi, L., Kovacs, G.T.A.: Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE Trans. Biomed. Eng. **53**(12), 2507–2515 (2006)
14. Lim, J.S.: Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system. IEEE Trans. Neural Netw. **20**(3), 522–527 (2009)
15. Homaeinezhad, M.R., Tavakkoli, E., Ghaffari, A.: Discrete wavelet-based fuzzy network architecture for ECG rhythm-type recognition: feature extraction and clustering-oriented tuning of fuzzy inference system. Int. J. Signal Proces. Image Process. and Pattern Recognit. **4**(3), 107–130 (2011)
16. Llamedo, M., Martinez, J.P.: Heartbeat classification using feature selection driven by database generalization criteria. IEEE Trans. Biomed. Eng. **58**(3), 616–625 (2011)
17. Zhou, H., Hou, K.M., Ponsonnaille, J., Gineste, L., De Vaulx, C.: A real-time continuous cardiac arrhythmias detection system: RECAD. In: Proceedings of 27th Annual International Conference of the Engineering in Medicine and Biology Society. Shanghai, China, pp. 875–881, 17–18 Jan 2005

18. Hu, S., Wei, H., Chen, Y., Tan, J.: A real-time cardiac arrhythmia classification system with wearable sensor networks. Sensors **12**(9), 12844–12869 (2012)
19. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation **101**(23), E215–220 (2000)
20. Moody, G.B., Mark, R.G.: The impact of the MIT-BIH arrhythmia database. IEEE Eng. Med. Biol. Mag. **20**(3), 45–50 (2001)
21. Berger, R.D.: QT variability. J. Electrocardiol. **36**(Suppl), 83–87 (2003)
22. Li, P., Liu, C.Y., Liu, C.C., Sun, H., Yang, J., Ma, G.Q.: Higher order spectra for heart rate variability and QT interval variability analysis: a comparison between heart failure and normal control groups. In: Computing in Cardiology, Hangzhou, China, pp. 309–312, 18–21 Sep 2011
23. Goh, K.W., Lavanya, J., Kim, Y., Tan, E.K., Soh, C.B.: A PDA-based ECG beat detector for home cardiac care. In: Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society. Shanghai, China, pp. 375–378, 17–18 Jan 2005
24. Tabakov, S., Iliev, I., Krasteva, V.: Online digital filter and QRS detector applicable in low resource ECG monitoring systems. Ann. Biomed. Eng. **36**(11), 1805–1815 (2008)
25. Li, P., Liu, C.C., Zhang, M., Che, W.B., Li, J.: A real-time QRS complex detection method. Acta Biophys. Sin. **27**(3), 222–230 (2011)