Providing an Efficient Algorithm for Finding R Peaks in ECG Signals and Detecting Ventricular Abnormalities With Morphological Features

Abstract

Ventricular arrhythmias are one of the most important causes of annual deaths in the world, which may lead to sudden cardiac deaths. Accurate and early diagnosis of ventricular arrhythmias in heart diseases is essential for preventing mortality in cardiac patients. Ventricular activity on the electrocardiogram (ECG) signal is in the interval from the beginning of QRS complex to T wave end. Variations in the ECG signal and its features may indicate heart condition of patients. The first step to extract features of ECG in time domain is finding R peaks. In this paper, a combination of two algorithms of Pan–Tompkins and state logic machine has been used to find R peaks in heart signals for normal sinus signals and ventricular abnormalities. Then, a healthy or sick beat may be realized by comparing the difference between R peaks obtained from two algorithms in each beat. The morphological features of the ECG signal in the range of QRS complex are evaluated. Ventricular tachycardia (VT), ventricular flutter (VFL), ventricular fibrillation (VFI), ventricular escape beat (VEB), and premature ventricular contractions (PVCs) are abnormalities studied in this paper. In the classification step, the support vector machine (SVM) classifier with Gaussian kernel (one in front of everyone) is used. Accuracy percentages of ventricular abnormalities mentioned above and normal sinus rhythm are respectively obtained as 95.8%, 92.8%, 94.5, 98.9%, 91.5%, and 100%. The database of this paper has been taken from normal sinus rhythm and MIT-SCD banks available on Physionet.org.

Keywords: *Electrocardiography, heart conduction system, humans, Pan–Tompkins algorithm, state logic machine, support vector machine, ventricular flutter, ventricular premature complexes*

Mohammad Pooyan,
Fateme Akhoondi[1]

Biomedical Engineering Department, [1]Department of Engineering, Shahed University, Tehran, Iran

Introduction

Analysis of electrocardiogram (ECG) signals is so important in detecting heart diseases, because such diseases may cause sudden deaths. Ventricular abnormalities are one of the most dangerous heart diseases. Accurate and early diagnosis of these disorders is crucial in patients with ventricular abnormalities. The electrical activity of the ventricles on ECG is in the range of QRS complex to T wave end. One group of ventricular abnormalities is ventricular arrhythmias. Ventricular arrhythmias studied in this article are ventricular fibrillation (VFI), ventricular tachycardia (VT), and ventricular flutter (VFL). Premature ventricular contractions (PVCs) and ventricular escape beat (VEB) are two ventricular abnormalities studied in this paper.

Analysis and classification of ECG beats have been done using numerous methods.

Inan *et al*.[1] first separated PVC beats from other healthy and cardiac abnormal beats by using wavelet transform and then classified them using time domain features and artificial neural networks. An overall accuracy of 95.6% was reported in this article. In Martis *et al*.,[2] the PCA has been used as a tool for ECG beat classification (normal bits, LBBB, PVC, RBBB). In this article, three features extraction methods [ECG rate basic elements, the basic elements of the linear prediction error signals and the basic elements of discrete wavelet transform (DWT) coefficients] were compared. In Ebrahimzadeh and Khazaee,[3] the ECG signal has been investigated using wavelet coefficients approximation (using three long features) and RBF neural network as a classifier. In this paper, five types of ECG beats (normal, LBBB, RBBB, PVC, and APC) are classified. In Korürek and Doğan,[4] the authors used particle Swann optimization

Address for correspondence:
Dr. Mohammad Pooyan, Engineering Faculty, Biomedical Engineering Department, Shahed University, Tehran, Iran.
E-mail: pooyan@shahed.ac.ir

Website: www.jmss.mui.ac.ir

and radial basis function neural network (RBFNN) for classifying six types of ECG beats. In de Chazal *et al.*,[5] the authors classified four types of ECG beats (normal beat, ventricular ectopic beat (VEB), supraventricular ectopic beat (SVEB), and fusion of normal and VEB). In this paper, the classification is performed using time domain features and linear kernel. In Hosseini *et al.*,[6] the combination of independent features and compressed ECG data is used as an input to the multilayered perceptron network. The accuracy of 88.3% has been reported.

In Thomas and Das,[7] the classification of five types of ECG beats (NB, Paced Beat, LBBB, RBBB, and PVC) is performed using wavelet coefficients in fourth and fifth scales and four features from QRS complex are considered in every cycle. The performance of the proposed method of this paper is compared with extracted statistical features using DWT. The overall accuracy of 97.8% has been reported.

In these abnormalities, we faced the dilemma of finding the location of R peaks, and no common method has proper functionality in finding R peaks. In this study, an algorithm has been provided to solve this problem.

The data used in this study are as follows:
(1) Normal sinus database (containing five 1-min signals or 350 beats) with frequency of 128 Hz.
(2) Sudden Cardiac Death Database (consisting of three 1-h signals) with frequency of 256 Hz.

In this study, the dataset is extracted from the website Physionet.org. Investigations in this paper are on Lead II. Table 1 reports the number of abnormal beats which are discussed in this study.

Materials and Methods

This part of the article consists of two parts: how to detect R peaks, and feature selection, which will be explained in the following.

QRS detection

To improve R peaks diagnosis, plenty of methods have been used. Many methods such as derivative method,[8] Hamilton–Tompkins algorithm,[9] wavelet transform

Table 1: Numbers of beats for each class

Class	Beat numbers
VT	80 beats
VFL	70 beats
VF	90 beats
PVC	60 beats
VSE	27 beats
NSR	350 beats

method, and Hilbert transform (HT) method are used to detect R peaks.

Some methods were applied for QRS detection by some translations, and more complex methods did not use from time dimension directly; some of them were mentioned as following.

In Rodríguez *et al.*,[10] the authors were given new approach for the detection of QRS complex in different arrhythmias. QRS was detected using HT and the adaptive threshold. Features were extracted using the principal component analysis.

In Biswas *et al.*,[11] the authors proposed peak synchronization that is novel method and it measures the simultaneity of occurrence of peaks in the signals.

In Ricardo Ferro *et al.*,[12] the authors employed processing of original signal, which included Shannon energy envelope (SEE) estimator, HT, and moving average (MA) filter. Data consisted of 10 records of 5 min length and different signal-to-noise ratios (15, 12, and 9 dB). The accuracy was applied to 100%.

In paper,[13] the authors used a combination of some signal processing techniques, which were Hilbert, wavelet transforms, and an adaptive thresholding method. This combination was named WHAT.

A common method for R peaks detection (such as Pan–Tompkins[14]) does not function properly through cardiac abnormalities occurrence. This study suggests a method based on a combination of Pan–Tompkins and state logic machine which has improved R peaks detection through abnormalities. State logic machine can remove high and low frequency noises.[1] Both algorithms detect the same location for R peaks in the ECG signal containing normal sinus rhythm and the signals not having acute abnormalities. The difference between R peak location detection via Pan–Tompkins algorithm and state logic machine in every single beat for normal sinus signals shows no significant difference or a maximum of 10_samples difference. However, the locations of R peaks obtained from the Pan–Tompkins algorithm and state logic machine are different in beats in the ECG signal having acute ventricular abnormalities. R peaks difference obtained from the two algorithms in two patient (PVC and VFL) and healthy cases are presented in Figures 1 and 2.

Moreover, a template of normal sinus beat is used in the algorithm. For this beat, the autocorrelation in zero lag is calculated (which is indicated by $a1$). This value is used in steps of thresholding and comparison. Then, the cross-correlation is calculated at zero lag between each beat and normal sinus beat template. Two conditions to classify data are studied in this paper. If both conditions are present in the meantime, that beat would be a normal sinus beat.
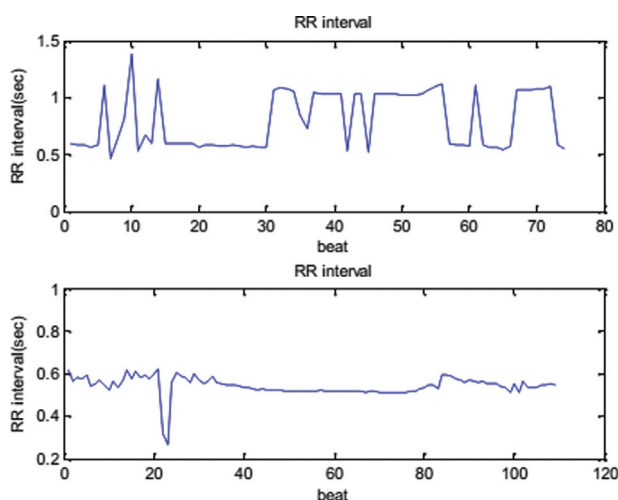
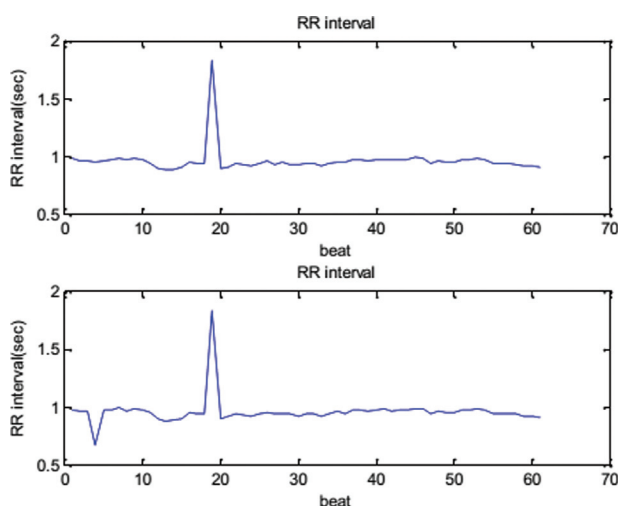Figure 1: RR intervals obtained from Pan–Tompkins and state logic machine, respectively, for a patient

Figure 2: RR intervals obtained from Pan–Tompkins and state logic machine, respectively, for healthy case
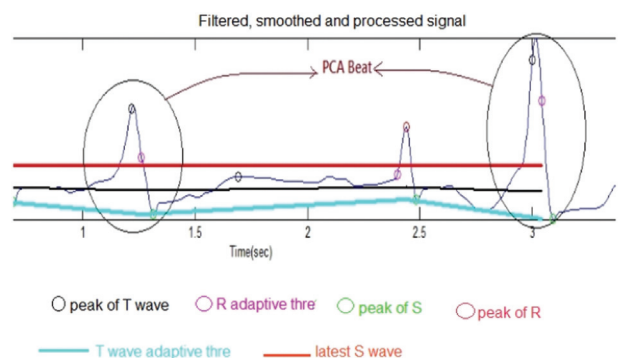

Figure 3: R peaks in PVC beats was detected wrongly

The first condition is, when the cross-correlation between the normal sinus beat template and studied beats from signals in zero lag (which is indicated by $a2$) is more than 0.3 of $a1$ and the difference of R peaks obtained from two algorithms (Pan–Tompkins and state logic machine) is smaller than

Table 2: FP, FN, TP and TN of R peak detection in ventricular abnormalities

Abnormality	FP	FN	TP	TN
PVC	26%	43%	57%	74%
VT	41.3%	48%	52%	58.7%
VFL	31.9%	46%	54%	68.1%
VFI	71%	62%	38%	29%

10, this beat is normal; otherwise, this beat is abnormal. The coefficient (0.3) is obtained from the average of the division $a2$ to $a1$. The beats' labels are available in Physionet. org website.

A significant point is that not all the beats are sampled with the same frequency. Also, all beats are downsampled to the lowest sampling frequency (128 Hz). Sudden cardiac death database (containing three 1-h signals) with frequency of 256 Hz and normal sinus database (containing five 1-min signals) with frequency of 128 Hz are sampled. The second condition is that if $a2$ is less than 0.3 of $a1$ or if the difference of R peaks obtained from two algorithms in the studied beat is more than 10, these beats have acute ventricular abnormalities.

At intervals of some ventricular arrhythmias in the ECG signals (such as PVC), no algorithm can find R peaks correctly (as shown in Figure 3, R peaks in PVC beat was marked wrongly in seconds 1.3 and 3).

Then, to find R peaks locations in ventricular abnormal beats, 70% of the interval between two consequent R peaks obtained from state logic machine is considered and the maximum peaks and their locations are calculated. The differences between the amplitudes of previous and next samples of this sample (obtained from previous step) are calculated to insure that is a real peak. Then, an interval of 0.06 s is considered after and before this sample and the minimum amplitudes are considered as Q and S peaks, respectively. After that, the features associated to QRS complex are obtained. R peaks are obtained using this algorithm for PVC beats with the TP of 57% and FN of 43%, if none of the other algorithms (state logic machine and Pan–Tompkins) are able to find R peaks. FP, FN, TP, and TN of detected R peaks for ventricular abnormalities in ECGs studied in this paper are according to Table 2.

Studied morphological features are as the following: the amplitude of R peak, the rising and falling slopes of QRS complex, QS interval, positive and negative areas (the sign of this value is negative) of the complex QRS [Eqs. (1) and (2) and Figure 4].

$$\sum_{0}^{n} 0.5(s(t) + |s(t)|) \qquad (1)$$
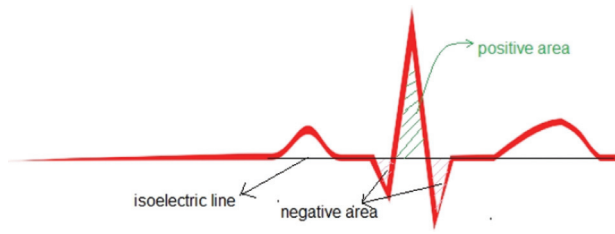
Figure 4: The positive and negative QRS complex area

$$\sum_{0}^{n} 0.5(s(t) - |s(t)|) \tag{2}$$

All of the features are obtained on Lead II.

State logic machine algorithm

In state logic machine algorithm, the ECG signal has passed through band pass filter with zero phase. This filter is necessary to remove background noise fluctuations and power line noise. Then, a 1_second window has been determined on the ECG signal and the average of this window has been calculated (m).

The state logic machine considered two conditions for each of the R and S and T waves.[15] Estimating the location is done by considering the first condition, and definite presence of R peaks in this interval is determined by considering the second condition and its constraints. Each condition has been different for peak detection in ECG at each step.

At zero state, the product of m and an initial weight ($w = 1.8$) is calculated. If this product is less than the mean value calculated in a 15_sample interval ($m1$) from the one second interval, and also if the amplitude of the studied sample is larger than the rest of the samples, the amplitude and the location of the first sample from the one second window is stored as R peak; then, we switch to state of one, which reports the definite presence of R peaks.[15] In summary [Eq. (3)]:

$$\text{if}(m1 > m * \text{weight} \subsetneq S(i) > S(j+1))$$

$$(j = 1:3; i \text{ in each step is fixed}) \text{then } S(i) \text{ is stored as R peak.} \tag{3}$$

In state of one, the stored amplitude from zero state is compared with the next four samples, to ensure that is the maximum one. Each sample in this window must be less than its previous sample. If such a condition is imposed, the index from zero state is stored as the R peak amplitude and its location. An interval containing 0.04 s before this location until the location of this sample is considered, and minimum amplitude is assumed as Q peak. Also, weight is updated in this state, which means that if the number of detected R peaks is more than eight, 0.3 of mean of these eight R peaks is

achieved and it is divided to m. then, the state will be equal to two.[15] In the state of two, if $m1$ is less than m, we switch to state of three for finding S peak. The period of investigation for state of three is 0.2 s. If, among the eight samples, the studied sample is less than the rest of the samples and if each sample is less than its previous sample, we ensure that the studied sample is S. Then, the state is equal to four.[15] In the state of four, if $m1$ is less than m, we switch to state of six for finding T peaks.

In state of six, the studied interval of 0.7 s is assumed. The threshold is considered based on the recent values of S and m. If $m1$ is more than this threshold [Eq. (4)], this condition is acceptable for three consecutive samples, and the peak of the studied sample is more than its subsequent eight samples (each sample in this eight-sample window should be smaller than the previous sample), this sample is stored as T peak and state is updated to six.

$$\text{threshold} =$$
$$|(|\text{mean of one second of signal}| - |S_{\text{amplitude}}(\text{end})|)| * \frac{3}{4}$$
$$+ S_{\text{amplitude}}(\text{end}) \tag{4}$$

State of six is for prevention of finding several R peaks in a 0.4 s window. After this delay, the algorithm switches to zero state and this algorithm is repeated for each 1_second window.[15]

The flowchart of the state logic machine is shown in Figure 5.

Classification

After extracting signal features, support vector machine (SVM) by (Gaussian) radial basis function kernel is used for classification of ventricular abnormalities. This is done for obtaining more reliable responses. The method used in SVM is one-against-all.[16,17] To classify each arrhythmia, data segmentation includes 8% of training data and 20% of testing data. The accuracy of the data is obtained by six times shuffles.

In this research, eight records of ECG signals were used from physionet. These signals include sudden cardiac death records for three patients and five normal sinus rhythms of five healthy persons. The feature extraction step extracted six features from more than 300 beats of MIT-SCD that included six acute abnormalities studied in this research, and 350 beats from normal sinus rhythm. These features, including the amplitude of R peak, the rising and falling slope of QRS complex, QS interval, and positive and negative areas of the complex QRS, were obtained from each beat. The data with more details were shown in Table 1.

The classification accuracies of normal sinus rhythm, VT, VFL, VFI, PVC, and VEB rhythm were evaluated. The accuracies achieved from this essay are compared with previous studies and the maximum accuracy is reported.
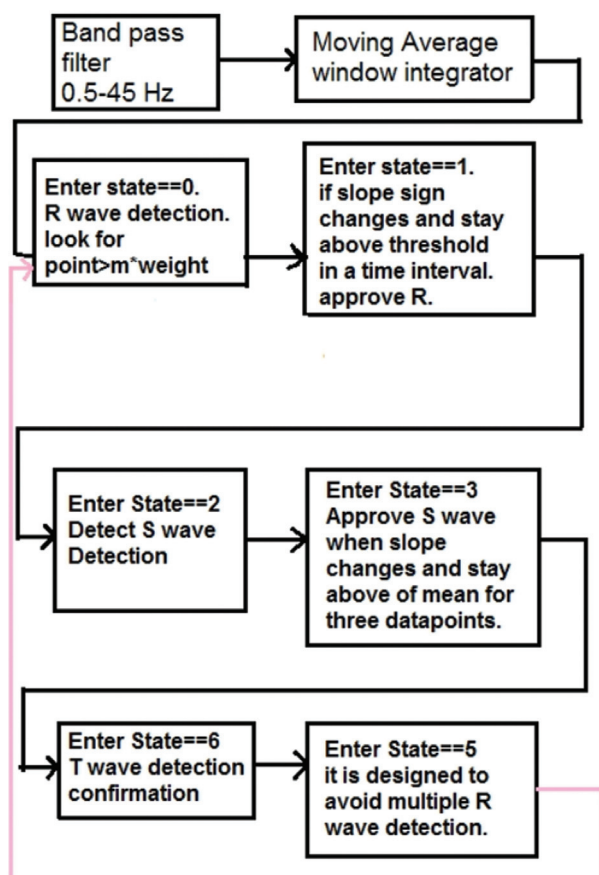
Figure 5: Flowchart of state logic machine

Table 3: Percent of accuracies for studied ventricular abnormalities in this paper

Name of class	Accuracy from this study	Maximum accuracy in previous study
VT	95.8	92.75[18]
VFL	92.8	91.36[18]
VFI	94.5	88.27[18]
PVC	98.9	96.66[19]
VES	91.5	51[20]
NSR	100	100[20]

By using the combined algorithm in this paper, we obtained acceptable accuracy percentages compared to the previous studies. Percent accuracies have been reported in Table 3.

Discussion and Conclusion

The algorithm used in this study was applied to increase the classification reliability of normal and abnormal beats in ECG real time signals and a combination of two algorithms (Pan–Tompkins and state logic machine) was used for finding R peaks in beats containing acute ventricular abnormalities.

In this research, morphological features were applied for classification of the aforementioned ventricular abnormalities. In this research, studying beat was done in beat-to-beat, real time for classification of ventricular arrhythmias. As a result, studying beat-to-beat may help the medicines for faster recognition of abnormalities rather than heart rate.

If each abnormality is studied at each lead and the most effective features in each lead are obtained, the abnormality classification will have very high accuracy. Independent features of leads will be very useful if they are diagnosed with high accuracy.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Inan OT, Giovangrandi L, Kovacs GT. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE Trans Biomed Eng 2006;53:2507-15.
2. Martis RJ, Acharya UR, Mandana KM, Ray AK, Chakraborty C. Application of principle component analysis to ECG signals for automated diagnosis of cardiac health. Expert Syst Appl 2012;39:11792-800.
3. Ebrahimzadeh A, Khazaee A. An efficient technique for classification of electrocardiogram signals. Adv Electric Comput Eng 2009;9:89-93.
4. Korürek M, Doğan B3 ECG beat classification using particle swarm optimization and radial basis function neural network. Expert Syst Appl 2010;37:7563-9.
5. de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng 2004;51:1196-206.
6. Hosseini HG, Reynolds KJ, Powers D. A multi-stage neural network classifier for ECG events. IEEE EMBS 2001;2:1672-5.
7. Thomas M, Das MK. Classification of cardiac arrhythmias based on dual tree complex wavelet transform. International Conference on Communication and Signal Processing, 2014. p. 3-5.
8. Reddy DC. Biomedical Signal Processing: Principle and Techniques. New Delhi: TMH Publication; 2005. p. 254-300.
9. Hamilton PS, Tompkins WJ. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. IEEE Trans Eng Biomed Eng 1986;33:1157-65.
10. Rodríguez R, Mexicano A, Bila J, Cervantes S, Ponce R. Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis. J Appl Res Technol 2015;13:261-9.
11. Biswas R, Khamaru K, Majumdar KK. A peak synchronization measure for multiple signals. IEEE Trans Signal Process 2014;62:4390-8.
12. Ricardo Ferro BT, Ramírez Aguilera A, Fernández de la Vara Prieto RR. Automated detection of the onset and systolic peak in the pulse wave using Hilbert transform. Biomed Signal Process Control 2015;20:78-84.
13. Rabbani H, Mahjoob MP, Farahabadi E, Farahabadi A. R peak detection in electrocardiogram signal based on an optimal

combination of wavelet transform, Hilbert transform, and adaptive thresholding. J Med Signals Sens 2011;1:91-8.

14. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985;32:230-6.

15. Segghamiz H. ECG Q R S Wave Online Detector; 2014. Available from: http://www.mathworks.com/matlabcentral/fileexchange/45404-ecg-q-r-s-wave-online-detector. [Last accessed 2016 Jan 12].

16. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 1992. p. 144-52.

17. Vapnik V. The Nature of Statistical Learning Theory. Springer Science & Business Media; 2013.

18. Lee SH, Ko HC, Yoon YR. Classification of ventricular arrhythmia using a support vector machine based on morphological features. Annual International Conference of the IEEE EMBS, 2013. p. 5785-8.

19. Ahmed R, Arafat S. Cardiac arrhythmia classification using hierarchical classification model. International Conference on CSIT, 2014. p. 203-7.

20. Matul Imah E, Jatmiko W, Basaruddin T. Adaptive multilayer generalized learning vector quantization (AMGLVQ) as new algorithm with integrating feature extraction and classification for arrhythmia heartbeats classification. 35th Annual International Conference on Systems Man and Cybernetics, 2012. p. 150-5.