BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 6, No. 2, April 2018

Classification of PVC Beat in ECG Using Basic Temporal Features

Y. Kaya

Abstract—**Premature ventricular contraction (PVC) is one of the most important arrhythmias among the various hearth abnormalities. Premature depolarization of the myocardium in the ventricular region causes PVC and it is usually associated with structural heart conditions. Arrhythmias can be detected by examining the ECG signal and this review requires large-size data to be examined by physicians. The time spent by the physician in examining the signal can be reduced using CAD systems. In this study, we propose a high performance PVC detection system using the feature extraction and classification scheme bringing low computational burden. The test set consisting of 81844 beats from the MIT-BIH arrhythmia database was used for the experimental results. We compared the performances of the various classifiers using proposed feature set in the experiments and obtained classification accuracy of 98.71% using NN classifier.**

Index Terms—**Arrhythmia, Classification, Decision Tree, Heartbeat, k-Nearest Neighbor, k-NN, Neural Network, Premature Ventricular Contraction, PVC, Support Vector Machine.**

I. INTRODUCTION

AN ECG IS A SIGNAL that can be easily obtained with electrodes placed on the human body, containing important information indicating the abnormal state of the cardiovascular system. Detection of different types of heartbeats is vital to identify cardiac disorders. Premature ventricular contraction (PVC) is one of the most important arrhythmias among the various anomalies related to cardiac rhythms [1], [2]. Premature depolarization of the myocardium in the ventricular region causes PVC and it is a common arrhythmia usually found in adults. It is estimated to have a prevalence of between 1% and 4% of the general population and usually associated with structural heart conditions and increases the risk of sudden death [2].

ECG signals need to be analyzed to detect arrhythmias, and this analysis is a time-consuming process that requires cardiologists to examine large-scale data. The accurate and rapid detection of PVC in the ECG signal is closely related to the correct identification of the features to represent a heartbeat. When PVC beat is examined, it is quite easy to distinguish it from normal sinus rhythm.

Y. KAYA, is with Department of Computer Engineering of Karadeniz Technical University, Trabzon, Turkey, (e-mail: yasin@ktu.edu.tr.).
Manuscript received August 9, 2017; accepted Nov 16, 2017.
DOI: 10.17694/bajece.419541

The signals can accurately and quickly be distinguished between large amounts of data by using computer aided diagnosis (CAD) systems and the time spent by the physician in examining the signal can be reduced. At this point, it is important to develop methods to distinguish between PVC and normal beats.

Researchers have done a lot of work for detecting of the PVC. The authors have proposed different models for feature extraction, noise elimination, feature size reduction and classification schemes for classification of arrhythmias in the following articles.

Liu et al. proposed a deep learning method for the recognition of PVC in children [3]. Kaya and Pehlivan compared various methods to classify PVC and investigated the model that gives the best result [2], [4]. Zhou et al. used deep neural network (NN) and rule inference to detect PVC [5]. Xiuling et al. proposed a model containing Lyapunov exponents and LVQ neural network to classify PVC beats [6].

Bortolan et al. compared the classification capabilities and the learning capacities of the k-nearest neighbor (k-NN), NN, fuzzy logic (FL) and, discriminant analysis (DA) classifiers for distinguish normal and PVC beats. The authors used 26 shape features in their work, which consisted of amplitude information, area, special interval times and QRS metrics. k-NN classifier reported to be more effective than other classifiers [7]. Ebrahimzadeh and Khazaee proposed a method to distinguish the PVC beat from normal and other beats. The authors used wavelet transform to eliminate noise in the ECG signal. They used one temporal and 10 morphologic features [8]. In another study, Christov et al. used both leads from the ECG signals from the MIT-BIH arrhythmia database to extract the feature for classification of PVC beats. They used k-NN as the classifier in the study and achieved the classification accuracy of 96.7% [9]. Jenny et al. used discrete wavelet transform to reduce noise in the signal, independent component analysis for dimension reduction and k-means and fuzzy c-means to classify for PVC beats [10].

In recent years, researchers have proposed new studies on wearable ECG analysis systems and mobile ECG analysis systems. One of the most important constraints for the systems is the calculation load. For this reason, it is very important for new methods developed to bring a low calculation burden [11].

In this study, we propose an approach for the classification of normal (N) and PVC beats. The main purpose of the work is to realize a high performance PVC detection system using the feature extraction and classification scheme bringing low computational burden. Unlike other studies in the literature, we achieved high classification performance for the classification of PVC beats using three basic features. Extraction of the features from the signal did not cause calculation loads on the system. For the experimental tests, the test set consisting of 81844 beats from the MIT-BIH arrhythmia database [12] was used in the study. We used k-NN, NN, support vector machine (SVM) and, decision trees (DT) classifiers to calculate and compare experimental test results. We distinguished between PVC and N beats with 98.71% classification accuracy using the NN classifier.

II. Material and Method

A. ECG Database

MIT-BIH arrhythmia database consisted of 48 ECG records with two-channel. Each record contained about 30 minutes ECG data. All of these records were obtained from 47 patients examined by the BIH Arrhythmia laboratory between 1975 and 1979. The first section of the database, numbered 100-124, was generated from 23 randomly selected 24-hour records. The second part of the database was carefully selected by cardiologists and numbered 200-234 for important clinical events [12], [13]. Fig.1. shows the normal beat and three types of PVC beat.

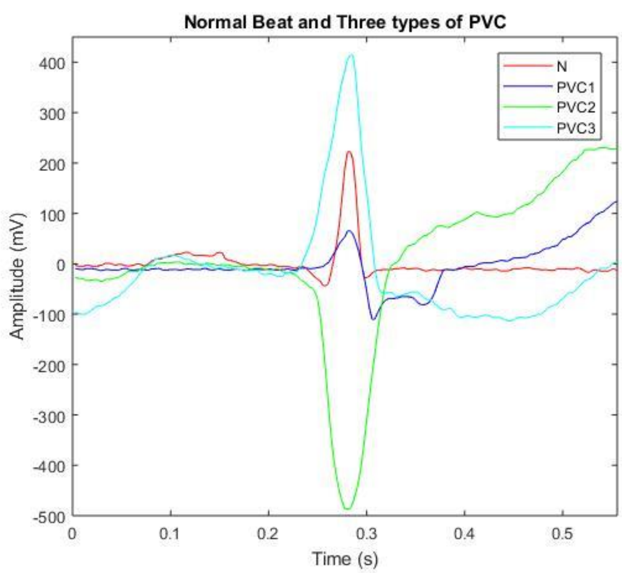

Fig.1. Normal beat and three types of PVC.

In the study, we used 46 ECG records containing MLII lead from the MIT-BIH arrhythmia database but two records without MLII were not preferred.

B. Classification of PVC beats

Decision-making in the classification of ECG beats is a three-step process similar to other machine learning methods.

Signal preprocessing: The fluctuations in the ECG signal are removed. In this step, the noise-free signal is split into beats and ready for feature extraction.

Feature extraction: The features to represent a beat are calculated using specified mathematical and statistical calculation methods. The method used at this step affects the accuracy of classification. The identification of better representing features of the beats provides the classification algorithm to learn better the data during the training step and gives better results in the test step.

Classification: In the classification step, tests are carried out with the proposed classification scheme. At this stage, a classification model must be determined according to the problem.

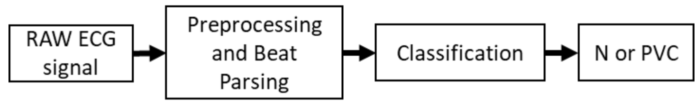

Fig.2. Proposed PVC detection approach

Fig.2. demonstrates the proposed approach for the detection of PVC beats. We determined three attributes in the feature calculation step in the study. These were the previous R-R interval (RRP), the next R-R interval (RRN), and arithmetic mean values of 50 amplitude values (MEAN) centered active R peak. The detection process successfully was performed with three features that can be obtained simply from the signal.

III. Experimental Results

We calculated the experimental results using the records from the MIT-BIH arrhythmia database in the study. Since the previous and next R-R values were calculated, the first and last beat in the signal files were not included in the analysis. Except for these beats, all PVC and N beats in the specified 46 signal files were included in the experiment set. These consisted of 7122 PVC beats and 74722 normal beats.

We used Matlab software to remove noise from the signal, beat parsing, and feature calculation steps. A beat parsing step was performed to obtain the signal values indicating a heart cycle. We calculated three features from these values, used as input vector in the classifiers, and evaluated classification performances.

A. Preprocessing and beat parsing

The signal was passed through various filters and the fluctuations found in the signal were reduced [2], [14], [15]. The most important of these fluctuations is the baseline wander that occurs with daily movements such of the patient as breathing, swelling, coughing, etc. during the ECG recording [15]. We removed the frequency components below 2 Hz from the signal using a high-pass filter to eliminate baseline wander [2].

B. Feature Calculation

We calculated the RRP, RRN and MEAN values used for the classification step for each beat in the feature calculation step. Fig.3. shows the calculation of RRP and RRN features. These values are the difference operation on the time axis.

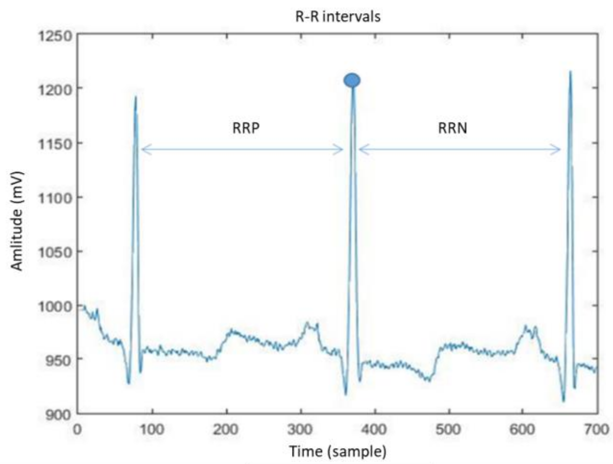

Fig.3. RRP and RRN features

Fig.4. shows the calculation of the MEAN feature. The window width was set to 50 for the calculation of the MEAN feature. We constructed a series of 50 sample amplitude values centered on R peak. Using these values, we calculated MEAN attribute. The notes in the database were used to determine the R peak. The arithmetic mean was calculated according to Equation (1) using the 50 amplitude values shown in Fig 4.

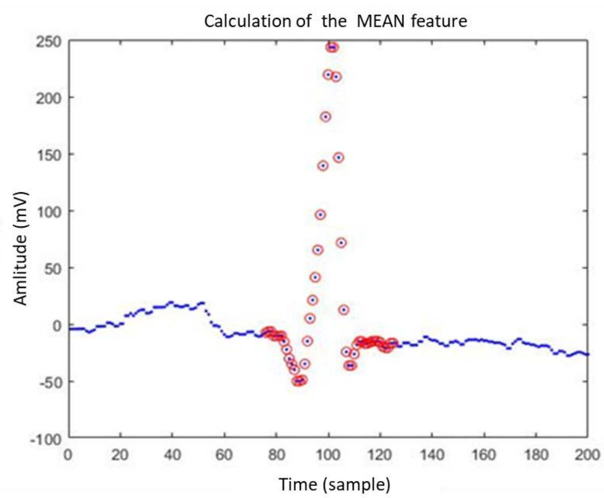

Fig.4. Calculation of MEAN feature

$$MEAN = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

where $x_i (\forall i \in 1..n)$ is the R peak centered amplitude values, n is the window width, which is 50.

C. Classification

We calculated three features in the previous step. In this step, the classification accuracy was calculated by classifying the features using k-NN, NN, SVM, and DT classifiers. The heartbeats shown in Table 1 were gathered from the MIT-BIH arrhythmia database. Normal beat was selected from 38 subjects and PVC beat from 35 subjects. We used 10-fold cross validation to evaluate results in the classification step. In 10-fold cross validation, the test data were divided into 10 separate pieces of equal width. One section used to test and the remaining nine were used for training the system at each step. After 10 repetitions, the overall performance of the system was calculated by evaluating the average of the classification performance achieved at each step.

We used classification accuracy, specificity, and sensitivity performance metrics to evaluate the results. Accuracy is defined as the ratio of the number of correctly classified samples to the total number of samples. Sensitivity is the ratio of the number of correctly classified positive samples to the total number of positive samples. Specificity is the ratio of the number of samples belonging to a correctly classified class to the total number of samples estimated for that class.

The accuracy, sensitivity and specificity measures are calculated from the confusion matrix using equations (2)-(4).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

where TP is correctly classified normal beat, TN is classified correctly PVC beat, FP is misclassified normal beat, FN is misclassified PVC beat.

TABLE I
HEART BEATS USED IN EXPERIMENTS AND RELATED RECORDS

Signal Files	N	PVC	Total
100	2237	1	2238
101	1858		1858
103	2080		2080
105	2524	41	2565
106	1505	520	2025
107		59	59
108	1737	17	1754
109		38	38
111		1	1
112	2535		2535
113	1787		1787
114	1818	43	1861
115	1951		1951
116	2300	109	2409
117	1532		1532
118		16	16
119	1541	444	1985
121	1859	1	1860
122	2474		2474
123	1513	3	1516
124		47	47
200	1742	825	2567
201	1623	198	1821
202	2059	19	2078
203	2527	444	2971

ISSN: 2147-284X

Signal Files	N	PVC	Total
205	2569	71	2640
207		105	105
208	1585	992	2577
209	2619	1	2620
210	2421	194	2615
212	922		922
213	2639	220	2859
214		256	256
215	3193	164	3357
217	244	162	406
219	2080	64	2144
220	1952		1952
221	2029	396	2425
222	2060		2060
223	2027	473	2500
228	1686	362	2048
230	2253	1	2254
231	314	2	316
233	2229	830	3059
234	2698	3	2701
Total	**74722**	**7122**	**81844**

IV. Discussions

In the study, 81844 samples of two classes from the MIT-BIH arrhythmia database were analyzed. In the experimental tests, almost all of the PVC and normal beats in the database were used. The arithmetic mean was calculated from 50 amplitude values of a beat and RRP and RRN temporal difference features were calculated for the classification process. These three features were fed to the k-NN, NN, SVM, and DT classifiers and the results were compared.

Table 2 shows the classification results. In the classification step, the k parameter for the k-NN classifier was set to one. A forward feed NN trained by the backpropagation algorithm was used. There was one hidden layer in the used architecture and the number of nodes in the hidden layer was set to 20 in the experimental tests. Another classifier used in experiments was SVM. In SVM, kernel type was defined as the radial based function (RBF) and gamma and C parameters were used as zero. DT was the last classifier used in experiments. We determined the maximum depth as 20, the minimum gain as 0.2, the minimum branch size as two, and minimum division number as four in the classification step with the KA. Experiments demonstrated that the best result was obtained with NN.

TABLE II
CLASSIFICATION RESULTS (%)

Method	Accuracy	Sensitivity [a]	Specificity [a]
k-NN	98.58	91.52	99.26
NN	98.71	90.80	99.46
SVM	98.55	87.67	99.58
DT	98.30	83.60	99.70

a. Specificity and Sensitivity values were calculated for the case where the PVC beat was positive class.

Table 3 shows the confusion matrix for the test performed with NN. In the study, high performance was achieved by using three simple time domain attributes. The fact that all PVC beats in the database were used in the experiments confirms the validity of the results obtained in the study.

TABLE III
CONFUSION MATRIX OF NN CLASSIFICATION

Class	True N	True PVC	Class Precision
Prediction N	74321	655	99.13%
Prediction PVC	401	6467	94.16%
Class Recall	99.46%	90.80%	

The proposed method shows that high performance can be achieved by using only three features when compared with other studies in this topic. Christov et al. classified PVC beats and reported the classification performance of sensitivity of 96.9% and specificity of 96.7% [9]. Jenny et al. used an unsupervised learning method and therefore achieved a lower performance than the other recommended methods [10]. Similarly, in other study, the authors classified PVC beats using k-NN and obtained specificity of 98.7% and sensitivity of 91.3% [7].

Table 4 summarizes the methods, the number of features, and the classification performance achieved in the proposed approach and the related studies. In our previous work in the same topic, we used 200 amplitude values for the classification of PVC beats. These 200 data were reduced to lower numbers using mathematical models [2].

TABLE IV
COMPARISON WITH OTHER STUDIES

The Authors	Method	Feature Size	Classification Accuracy
Bortolan et al. [7]	k-NN	26	98.7% spe. 91.3% sen.*
Liu et al. [3]	1D CNN	486	83%
Zhou et al. [5]	CDNN	150	99.41%
Christov et al. [9]	k-NN	26	96.7% spe. 96.9% sen.*
Ebrahimzadeh et al. [8]	NN	11	95.37%
Jenny et al. [10]	Fuzzy C-means	-	80.94%
Kaya et al. [2]	k-NN	17	99.63%
Proposed Approach	**NN**	**3**	**98.71%**

* Some authors did not specified classification accuracy. For this reason, specificity (spe.) and sensitivity (sen.) metrics are shown in the table.

In a more recent study, Liu et al. proposed a deep learning based method for perceiving PVC beats in children. The authors recorded the test data themselves and obtained the correct classification accuracy of 83% using the recommended method [3]. In a similar study using combined deep neural networks and rules inference, the authors achieved a classification success of 99.41%. The authors used the experimental set consisting of 3194 PVC beats and 46329 normal beats to achieve this success [5].

Researchers have worked extensively on these issues and have proposed complex models. The proposed models bring the computational burden. Used methods for feature extraction in the studies summarized in Table 4 were complex and computationally expensive.

V. Conclusions

In this study, we propose a classification scheme based on NN classifier in order to obtain the highest performance with minimum account load. The model uses three attributes that are easy to calculate in time domain. In this respect, the study uses fewer features than other studies in the literature. The use of almost all N and PVC beats in the database confirms the validity of the results obtained by the proposed method.

Due to the small number of attributes and simple calculation, and the simplicity of the classification scheme used, the proposed method can be integrated with the mobile ECG recording and analysis applications to be developed. The proposed model can be used in wearable ECG systems because it can be classified with low complexity and high accuracy.

References

[1] G. K. Lee, K. W. Klarich, M. Grogan, and Y.-M. Cha, "Premature ventricular contraction-induced cardiomyopathy: a treatable condition.," *Circ. Arrhythm. Electrophysiol.*, vol. 5, no. 1, pp. 229–36, Feb. 2012.

[2] Y. Kaya and H. Pehlivan, "Classification of Premature Ventricular Contraction in ECG," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 7, pp. 34–40, 2015.

[3] Y. Liu, Y. Huang, J. Wang, L. Liu, and J. Luo, "Detecting Premature Ventricular Contraction in Children with Deep Learning," *J. Shanghai Jiaotong Univ.*, vol. 23, no. 1, pp. 66–73, Feb. 2018.

[4] Y. Kaya and H. Pehlivan, "Classification of Premature Ventricular Contraction Beat Using Basic Temporal Features," in *International Advanced Researches & Engineering Congress-2017*, 2017, pp. 1313–1318.

[5] F. Zhou, L. Jin, and J. Dong, "Premature ventricular contraction detection combining deep neural networks and rules inference," *Artif. Intell. Med.*, vol. 79, pp. 42–51, Jun. 2017.

[6] X. Liu, H. Du, G. Wang, S. Zhou, and H. Zhang, "Automatic diagnosis of premature ventricular contraction based on Lyapunov exponents and LVQ neural network.," *Comput. Methods Programs Biomed.*, vol. 122, no. 1, pp. 47–55, Oct. 2015.

[7] G. Bortolan, I. Jekova, and I. Christov, "Comparison of four methods for premature ventricular contraction and normal beat clustering," in *Computers in Cardiology*, 2005, vol. 32, pp. 921–924.

[8] A. Ebrahimzadeh and A. Khazaee, "Detection of premature ventricular contractions using MLP neural networks: A comparative study," *Meas. J. Int. Meas. Confed.*, vol. 43, pp. 103–112, 2010.

[9] I. Christov, I. Jekova, and G. Bortolan, "Premature ventricular contraction classification by the K th nearest-neighbours rule," *Physiol. Meas.*, vol. 26, no. 1, pp. 123–130, Feb. 2005.

[10] N. Z. N. Jenny, O. Faust, and W. Yu, "Automated Classification of Normal and Premature Ventricular Contractions in Electrocardiogram Signals," *J. Med. Imaging Heal. Informatics*, vol. 4, no. 6, pp. 886–892, Dec. 2014.

[11] M. M. Baig, H. Gholamhosseini, and M. J. Connolly, "A comprehensive survey of wearable and wireless ECG monitoring systems for older adults," *Med. Biol. Eng. Comput.*, vol. 51, no. 5, pp. 485–495, May 2013.

[12] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database.," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[13] G. Moody and R. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," in *[1990] Proceedings Computers in Cardiology*, 1990, pp. 185–188.

[14] Y. Kaya and H. Pehlivan, "Comparison of classification algorithms in classification of ECG beats by time series," in *23nd Signal Processing and Communications Applications Conference (SIU)*, 2015, pp. 407–410.

[15] Y. Kaya, H. Pehlivan, and M. E. Tenekeci, "Effective ECG beat classification using higher order statistic features and genetic feature selection," *Biomed. Res.*, vol. 28, no. 17, pp. 7594–7603, 2017.

BIOGRAPHIES

YASIN KAYA was born in Adana, in 1979. He received the B.S. degree in statistics and computer science from the Karadeniz Technical University, in 1999. He received M.S. and Ph. D. degrees in Computer Engineering from the Karadeniz Technical University, Trabzon, in 2006 and 2017, respectively.

From 1999 to 2015, he was a lecturer with the Department of Informatics. Since 2015, he has been a Lecturer with the Distance Learning Application and Research Center, Karadeniz Technical University. He is the author of four books, three international indexed (SCIe, ESCI) articles, and eight international conference proceedings. His research interests include biomedical signal processing, arrhythmia detection, pattern recognition, and image processing.

