Classification of cardiac arrhythmia with respect to ECG and HRV signal by genetic programming

Article · January 2012

CITATIONS
12

READS
885

3 authors, including:

Mohammad Mehdi Ebadzadeh
Amirkabir University of Technology
111 PUBLICATIONS **1,380** CITATIONS

SEE PROFILE

Classification of cardiac arrhythmia with respect to ECG and HRV signal by genetic programming

Masih Tavassoli, Mohammad Mehdi Ebadzadeh, Hamed Malek

Abstract — Consistent or periodical heart rhythm disorders may result cardiac arrhythmias. In this article, heart rate variability (HRV) signals are analyzed and various features including time domain, frequency domain and nonlinear parameters are extracted. Moreover, additional nonlinear features are extracted from electrocardiogram (ECG) signals. These features are helpful in classifying cardiac arrhythmias. In this paper, genetic programming is applied to classify heart arrhythmias using both HRV and ECG features. Genetic programming selects effective features, and then finds the most suitable trees to distinguish between different types of arrhythmia. By considering the variety of extracted parameters from ECG and HRV signals, genetic programming can precisely differentiate various arrhythmias. The performance of proposed algorithm is evaluated on MIT–BIH Database. The results show that seven different types of arrhythmia classes including normal beat , left bundle branch block beat, right bundle branch beat, premature ventricular contraction, fusion of ventricular and normal beat, atrial premature contraction and paced beat are classified with an accuracy of 98.75%, 98.93% , 99.10%, 99.46%, 99.82%, 99.46% and 99.82% respectively..

Key Words — Arrhythmia, Electrocardiogram (ECG), Heart Rate Variability (HRV), Genetic Programming (GP), Feature Selection

I. INTRODUCTION

Heart is a muscular organ which is responsible to pump oxygenated blood throughout blood vessels by rhythmic contractions. Any disturbance in the heart rhythm can be very dangerous. Although cardiac arrhythmia is one of the leading causes of death, it can be treated if detected on time. Heart arrhythmia can cause too slow or too fast performance of the heart. To detect it, ECG and HRV signals are widely used. ECG signal records electrical performance of the heart. It contains a lot of important information related to the condition of the heart and one of the most important tools in detecting heart diseases. A typical ECG signal consists of the P-wave, QRS complex, and T-waves. The P wave is the result of slow-moving depolarization of the atria. QRS complex which is made of Q, R and S waves shows ventricular depolarization. The T wave represents repolarization of the ventricles, and is longer in duration than depolarization. HRV signal describes the variations between consecutive heartbeats. It is a non-offensive evaluation method of the nervous system which controls cardiovascular system and is a measurement of the interaction between sympathetic and parasympathetic activity in autonomic nervous system. HRV signal is a non-stationary signal and its changes can be interpreted as a current or upcoming disease.

Several methods for automatic detection and classification of cardiac arrhythmias have been proposed in literature, including: artificial immune recognition system with fuzzy weighted [1], threshold-crossing intervals [2], neural networks [3], fuzzy neural networks [8], fuzzy equivalence relations [12], Bayesian classifiers [16], support vector machines [17,22], wavelet transforms [18-20], combined wavelet transformation and radial basis neural networks [21], fuzzy logic combined with the Markov models [23] and the rule-based algorithms [24]. Some papers used techniques which are based on ECG segment [2,9,11,21,22,23,25]. In these papers, the various features of the ECG signal including the morphological features are extracted and used for classification of the cardiac arrhythmias. This is a time consuming procedure and the results are very sensitive to the amount of noise. An alternative approach would be to extract the HRV signal from the ECG signal [12-18,26,29] first by recording the R-R time intervals and then processing the HRV signal instead. This is a more robust method since the R-R time intervals are less affected by noise. One drawback of the proposed HRV-based algorithm is that some of the arrhythmia types such as the left bundle branch block and the right bundle branch block beats cannot be detected using only the heart rate variability features.

In this paper a new arrhythmia classification algorithm is proposed which is able to effectively classify seven types of arrhythmia. These arrhythmias are namely the normal beat (NB), left bundle branch block beat (LBBB), right bundle branch beat (RBBB), premature ventricular contraction (PVC), fusion of ventricular and normal beat (FUSION), atrial premature contraction (APC) and paced beat (PACE). In this article, various features from both ECG and HRV signal are extracted and given to a genetic programming to produce the suitable solution trees to distinguish between different types of arrhythmia. From the various identified features the proposed method selects the effective ones and categorizes the seven classes of heart arrhythmia highly precisely.

The reminder of this paper is formatted as follows: section 2 provides the overall block diagram of the proposed algorithm. In section 3, the database which is being utilized in this paper is introduced. Extracted features from HRV and ECG signal are explained in section 4. Genetic programming algorithm is introduced in the fifth section. The results are shown in section 6. Section 7 discusses about the result .Finally, Section 8 concludes the paper.

II. THE PROPOSED ALGORITHM

The block diagram of the proposed algorithm is demonstrated in Fig. 1. As seen, it consists of four steps: pre-processing on ECG signals to divide ECG signals to eight consecutive RR intervals and extract HRV signal, feature extraction from ECG and HRV signal, creating seven optimal trees to detect each arrhythmia by a genetic programming algorithm and finally arrhythmia classification. The following sections describe each block of this algorithm in more details.

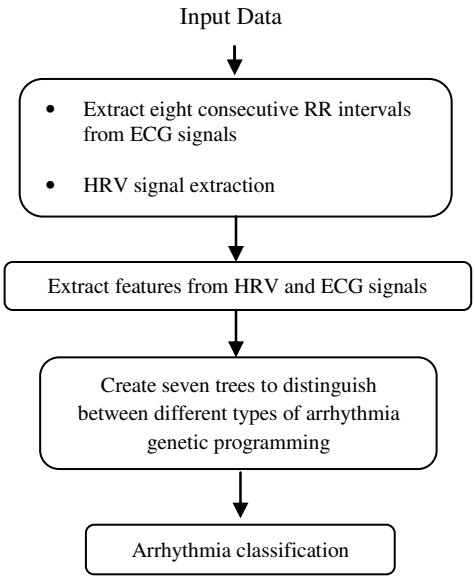

Fig. 1. Block diagram of the proposed arrhythmia classification algorithm.

III. DATABASE

In this paper, the proposed approach is tested using the MIT-BIH arrhythmia database. This database contains 48 ECG records. Each record is approximately 30 minutes long and it includes 109000 R-R intervals with sampling frequency of 360 Hz. Each beat has been annotated independently by two cardiologists. Their annotations were compared, consensus on disagreements was obtained, and the reference annotation files were prepared [27]. In most records, the upper signal is a modified limb lead II (ML II) and the lower signal is a modified lead V1 (VI). All of the ECG records of this article are chosen from lead II and include 8 sequential R-R intervals. Extracted records of this database include all seven classes of heart beat.

After dividing the ECG signals into 8 sequential RR intervals, HRV signal is extracted from calculating the time intervals between every two consequential R-wave in an ECG signal (R-wave is located in the maximum absolute value of the signal within the time window).

IV. FEATURE EXTRACTION

As seen in Fig. 1, features are extracted from both HRV and ECG signals. In the following subsections extracted features are explained.

A. Extracted features from HRV signal

Since the behavior of HRV signal includes both linear and non-linear behavior, a combination of these features is considered. These features include time domain, frequency domain and nonlinear parameters. Each ECG signal is divided into 8 consequential R-R intervals and each segment of HRV signal includes time distances between every two consequential R-wave in ECG signal.

1) Time domain features

Time-domain parameters of HRV are the easiest as they are based on common statistical methods. In this paper, seven commonly used time domain features are as follows: [28]:

- The mean value of the eight R-R intervals within each segment (Mean).
- The root mean square successive difference of the eight R-R intervals in each segment (RMSSD).
- The standard deviation of the 8 R-R intervals within each segment (SDNN).
- The standard deviation of differences between the adjacent R-R intervals within each segment (SDSD).
- The number of successive difference of intervals which differ by more than 50, 10 and 5 ms, respectively, divided by 8, the total number of the R-R Intervals within each segment (pNN50, pNN10, pNN5).

2) Frequency domain features

LF/HF: Although time domain features are important in classifying arrhythmia, they are not capable of distinction of sympathetic and parasympathetic content of the HRV signal [29]. For this purpose, HRV signal is transformed into frequency domain and the ratio of spectral power in lower bound (0.04-0.15Hz) to spectral power in upper bound (0.15-0.5Hz) is calculated. The lower bound frequency power is related to controlling temperature and cardiovascular mechanism and the upper frequency is related to the cardiac vagal activity.

3) Nonlinear features

The nonlinear properties of HRV can be analyzed using such as follow measures:

ApEn: Approximate entropy measures the complexity or irregularity of the signal. Large values of ApEn indicate high irregularity and smaller values of ApEn implies higher regularity [30]. The ApEn is computed as follows.

For each segment in HRV signal with length N, u_j is defined as follow:

$$u_j = (RR_j, \ldots, RR_{j+m-1}), j = 1, \ldots, N-m+1, \quad (1)$$

where m is called the embedding dimension and N is the number of measured RR intervals. The distance between these vectors is defined as the maximum absolute difference between the corresponding elements. For each u_j the relative number of vectors u_k for which $d(u_j, u_k) \le r$ is calculated by Eqs. (2) and (3)

$$d(u_j, u_k) = \max\{|RR_{j+n} - RR_{k+n}| \,|\, n = 0, \ldots, m-1\}, \quad (2)$$

$$C_j^m(r) = \frac{|\{u_k \,|\, d(u_j, u_k) \le r\}|}{N-m+1}. \quad (3)$$

Due to the normalization, the value of $C_j^m(r)$ is always smaller or equal to 1. Afterward, the mean of natural logarithm of each $C_j^m(r)$ over j is taken to yield:

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} \ln C_j^m(r). \quad (4)$$

Finally approximate entropy calculated by Eq. (5)

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r). \quad (5)$$

For calculating ApEn for each HRV segment, the value of m and r are chosen as $m = 2$ and $r = 0.2SDNN$.

SpEn: Spectral entropy evaluates the HRV signal complexity in frequency domain [31]. Shanon channel entropy estimates HRV entropy as

$$H = -\sum_f P_f \log(P_f), \quad (6)$$

where P_f is the value of the probability density function (PDF) of the process at frequency f.

Poincare´ plot: This plot is another technique for analysis of HRV signal [32]. It is a graphical representation of the correlation between successive R-R intervals. By considering Poincare´plot as a time series of RR_i, if each interval RR_{n+1} is plotted as a function of the previous interval RR_n, then the resulting plot is known as the Poincare´ plot. This plot shows the heart problem and some information about short term and long term oscillations. The Poincare´plot is derived by calculating *SD1* and *SD2* parameters which are standard deviations of RR_i distances from $y=x$ and $y=-x+2(RRm)$ lines respectively. *RRm* is the mean of RR_i. *SD1/SD2* describes relation between parameters. A common approach to parameterize the shape is to fit an ellipse to the plot as shown in Fig. 2.

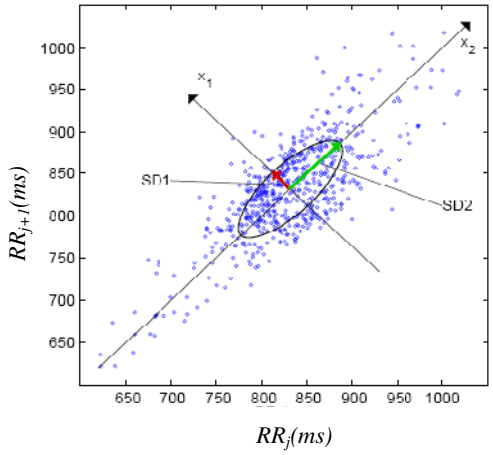

Fig. 2. Poincar´e plot analysis with the ellipse fitting procedure. *SD1* and *SD2* are the standard deviations in the directions *x1* and *x2*, where *x2* is the line-of-identity for which *RRj = RRj+1*.

Correlation Dimension: This parameter is an index of system complexity and indicates the number of independent variabilities to describe the behaviour of the system [33]. Correlation Dimension of chaotic system is always a fraction, but it can be either a fraction or an integer number for a random system. By considering HRV signal as a time series, the correlation dimension can be calculated as follows:

For each segment in HRV signal with length N, u_j is defined as follow:

$$u_j = (RR_j, \ldots, RR_{j+m-1}), j = 1, \ldots, N-m+1, \quad (7)$$

afterward by calculating $d(u_j, u_k)$ distances, the number of vectors shorter than r is calculated, as in Eqs. (8) and (9).

$$d(u_j, u_k) = \sqrt{\sum_{l=1}^{m} (u_j(l) - u_k(l))^2}, \quad (8)$$

$$C_j^m(r) = \frac{|\{u_k \,|\, d(u_j, u_k) \le r\}|}{N-m+1} \quad \forall k. \quad (9)$$

Then the mean value of $C_j^m(r)$ is calculated:

$$C^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} C_j^m(r). \quad (10)$$

The correlation dimension is calculated by Eq. (11)

$$D_2(m) = \lim_{r \to 0} \lim_{N \to \infty} \frac{\log C^m(r)}{\log r}. \quad (11)$$

As shown in Fig. 3, in practice, this limit value is approximated by slope of $\log C^m(r)$ versus $\log r$ when m is increased.

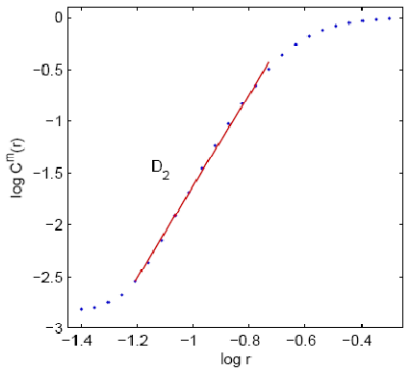

Fig. 3. Approximation of the correlation dimension D2 from the *log r*, *log* C^m *(r)* plot.

B. Extracted features from ECG signal

In this section, ECG signal is divided into 8 consequential RR intervals and by employing fractal dimension, lyapunov exponent and Hurst exponent, follow nonlinear features are extracted.

Fractal Dimension: The concept of fractal dimension that refers to a non-integer or fractional dimension originates from fractal geometry. This feature has been used in the analysis of ECG and EEG to identify and distinguish specific states of physiologic function [35]. Box-counting dimension method is used to calculate the fractal dimension. The fractal dimension of an ECG signal is based on chaos theory and is a quantitative measurement of the roughness of that signal. It can be used as an effective parameter in categorizing heart arrhythmias. To calculate fractal dimension, each dimension of the signal is divided to S segments and the box which contain ECG signals are counted. N is the number of such boxes.

$$dim_{box}(S) = \lim_{S \longrightarrow 0} \frac{\log N}{\log S}. \tag{12}$$

In practice, this limit value is approximated by slope of $\log N$ versus $\log S$ when m is decreased.

Lyapunov Exponent: In a dynamic system, the dependence on initial condition is described by Lyapunov Exponent (LE), and calculates the rate of deviating from roots. A positive LE shows that the distance of two points is increasing exponentially. This means that system tends to chaos. A negative LE indicates stable behaviour and a zero LE means that two close points on a root, keep their distance. By considering ECG signal as a time series of $X=x_0, x_1,\ldots,x_{n-1}$, it can be calculated as follows [34]:

The procedure is started at t_0 with x_0, next the point in the time series which is closest to x_0 is found and is called x_i. Then d_0 the absolute difference of these two points is calculated. dc_0 is the absolute difference of the two consecutive points of x_0 and x_i, namely x_1 and x_{i+1}. The same procedure is repeated for points x_1 to x_{n-1} and d_1 to d_{n-2} and dc_1 to dc_{n-2} are calculated. LE is approximately calculated by Eq. (13).

$$\lambda = \frac{1}{n-1}\sum_{t=0}^{n-1} \log \frac{d1_t}{d_t}. \tag{13}$$

Hurst Exponent: The Hurst exponent is based on Hurst investigations to detect the incoming water flow in dam that he built on the Nile River. The incoming water flow in dams was assumed to be random but Hurst found out a non-periodic cycles in incoming flows based on the previous data. Hurst test was over generalized to other phenomena which seem to be random but may have an organized pattern. The Hurst exponent is a measure of the smoothness of a fractal time-series based on the asymptotic behavior of the rescaled range of the process. The test procedure is as follows [34]:

By considering ECG signal as a time series of $U=u_0,u_1,\ldots,u_{T-1}$, divide this series to a consequential time series of length n into a contiguous subperiods. Each subperiod Ij is labeled with $j = 0,1,\ldots,a-1$ and each element in Ij is labeled $N[j][k]$ such that k = 0, 1,..., $n-1$. For each subperiod I_j, the mean value is calculated and data scale is normalized by Eqs. (14) and (15):

$$E_j = \frac{1}{n}\sum_{k=0}^{n-1} N[j][k], \tag{14}$$

$$X[j][k] = \sum_{i=0}^{k} (N[j][k] - E_j), \; k = 0,1,2,\ldots,n-1. \tag{15}$$

The sample standard deviation calculated for each subperiod I_j is:

$$R_{I_j} = max(X[j][k]) - min(X[j][k]). \quad 0 \le k \le n-1 \tag{16}$$

For each subperiod I_j, the value of R vector is calculated by Eq. (17).

$$S_{I_j} = \left(\frac{1}{n}\sum_{k=0}^{n-1} (N[j][k] - E_j)^2\right)^{1/2}. \tag{17}$$

The R_{Ij} is normalized with respect to a particular length n with Eq. (18).

$$(R/S)_n = \frac{1}{a}\sum_{j=0}^{a-1} \frac{R_{I_j}}{S_{I_j}}. \tag{18}$$

As a result amplitude R_{ij} is always non-negative. By increasing n such that T/n is always an integer, this procedure is repeated until $n=T/2$. Hurst offered these equations by using the half rule in statistics.

$$\log(R/S)_n = \log(c) + H \log(n), \tag{19}$$

in which R is the rescaled amplitude. S is the standard deviation of the time series, c is a constant, n is the length of the subperiod and the slope of the equation is the estimate of the Hurst exponent, H. According to Hurst results, if the

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.