MINING OF ECG SIGNAL FOR NEW DIAGNOSTIC INFORMATION

RAJIV KUMAR NATH

Department of Computer Science & Engineering, College of Engineering & Technology, IFTM Campus, Moradabad, Uttar Pradesh 244001, India

SANJAY NATH

Pharmacy Department, IITM College, Agra, Uttar Pradesh, India

Abstract

This paper investigates a technique, which extracts new features having potentiality for giving more discriminatory clues from simple ECG (Electrocardiogram) signals. The extracted feature parameters, Average RR Signal Morphology (ARSM) and RR Interval Frequency Histogram (RIFH) of an ECG signal from pre-selected data segment of the MIT-database shows that they may be utilized to detect some clinically challenging problems like Paroxysmal disease of Heart. For checking the efficiency of a Multidimensional- feature parameter like those mentioned above, we have employed a novel two dimensional display method. Overall, it was shown that the same technique may be employed for comparative study of new as well as old feature parameters extracted from biomedical images and signals.

Keywords: ECG signal; Average signal morphology; RR Interval Frequency Histogram; feature comparison; Cardiac arrhythmia**.**

1. Introduction

The human Body is dynamic self-controlled system whose stability (homeostasis) is ensured by simultaneous functioning of distributed physiological system, neuroregulation, blood circulation etc. Continued functioning of all these life supporting system is reflected in real time in a complex pattern of physical field and radiation coming from human body like infrared, microwave, optical, acoustic radiation, electric and magnetic fields and radiations and it also depends on the atmosphere that usually surround the human as mentioned by [Prokhorov et al., 2003]. [Bezruchko et al., 1986] states that "A technique for determination of character and intensity of interaction between the elements of complex systems based on reconstruction of model equations for phase dynamics is extended to the case of short and noisy time series. Corrections, which eliminate systematic errors of the estimates, and the expressions for confidence intervals are derived". [Surve et al.,2004] and [Traditional Tibetan Medicine] states that noninvasive recording of pressure pulse waveform from the radial artery can be used for obtaining valuable diagnostic information, by analyzing it for temporal characteristics, spectral characteristics, and its cross-correlation with phonocardiogram and ECG. The pulse waveform can be recorded using the transducer of an electronic stethoscope. But, there may exist likely a chance of noise, which may corrupt the characteristics of the signal. Ancient Indian study also showed the effectiveness of pulse sensing based diagnostics where the observational facts tell that this Radial Artery Pressure Signal has a good discriminatory power for many diseases". All these studies point to the fact that any apparently simple biomedical signal may be considered as the cumulative effect of a complex spatiotemporal event going on inside the human physique. To study such cumulative effect of any Biomedical Signal (BMS), an attempt has been in this article by studying feature-phenomenology of human system from the information that is embedded or hidden within the apparently simple ECG signal. [Lepage et al., 2001] worked on automatic analysis method of the P-wave, based on lead II of a 12 lead standard ECG, which will be applied to the detection of patients prone to arterial fibrillation (AF), which is considered as one of the most frequent arrhythmias. [Surekha, 1986] states that "an automatic analysis method of the Pwave, based on lead II of a 12 lead standard ECG, which will be applied to the detection of patients prone to atrial fibrillation (AF), is found as one of the most frequent arrhythmias". [Gholam et al, 1998] states that "a set of efficient technique to extract important feature from the ECG data applicable in automatic cardiac arrhythmia classification. The selected parameters are divided into two main categories namely morphological and statistical features". Frank G. Yanowitz, MD [8-9] has discussed in detail about the various diseases information from the ECG Signal, and tried to find some more possible diagnostics.

2. Methodology and Algorithm

An Attempt has been made to develop an algorithm through coding system with the help of Matlab 5.0 environment‟ software. The analysis is based on forty set of data which has been obtained through internet.
Major headings should be typeset in boldface with the first letter of important words capitalized.

2.1. *Preprocessing*

A classification system works in a number of stages. Firstly, data collection is done from physionet site. Secondly, preprocessing operations, which includes the conversion of data in text format followed by removal of noise/artifacts and other processing techniques? These are generally performed by simple low-pass filtering. In this article emphasis has been put to study the characteristics of the signal-pattern developed during an ECG test, for each individual RR block. This approach has novelty in the sense that once you can splitup the signals for each RR block you may use this raw-signal in Toto as the signal feature parameter for biometric type of diagnostics.

2.1.1. *Detection* of RR peaks

In our work lot of improve mental or fine tuning job can be done. In case of ARSM parameter if we take signals for the RR blocks of maximum frequency range it may give us more accurate result. This work can be used for other bio-signal like EEG and MEG also. It describes technique, which extracts important features from the ECG signal data in semi -automatic detection of RR interval of an ECG signal in which the user specifies the range. The RR peaks of an ECG signal are detected with of finding out the maximum value of amplitude wave in the ECG signal. This maximum value should correspond to the one pick value of the R wave in the ECG signal. After this we fix the lower and maximum range of the R amplitude of the ECG signal with the help of
data heuristics. So that all R waves of the ECG signal are should be present in the lower range and maximum range of the ECG signal. There is a possibility of getting more than one R peak in one cycle.

To eliminate this ambiguity we are using clustering method to find out only one R value for one cycle.

2.2 *Extraction of features*

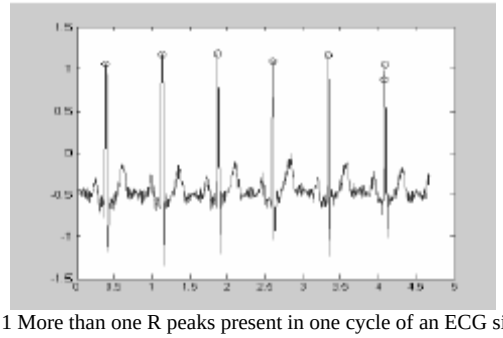


Fig. 1 More than one R peaks present in one cycle of an ECG signal

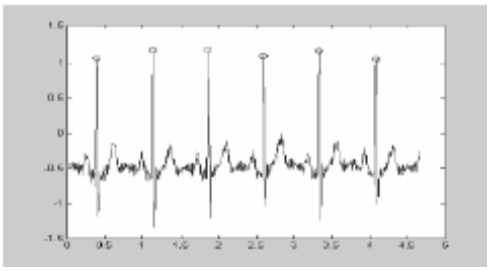


Fig. 2 Only one R peak present in one cycle of an ECG signal after preprocessing

In the feature extraction phase, we tried to extract some new but simple features to identify clinically challenging problem like paroxysmal of heart through the following methodologies:

2.2.1 *RR Interval Frequency Histogram*

The following steps adopted for the extraction of RR interval frequency histogram.
1) The total range of frequency histogram has chosen from zero to three seconds.
2) The whole range divided into 30 equal ranges and the frequency of RR interval for each of this 30 range calculated. This set of frequencies at different time interval range gives the RIFH parameter.

2.2.2 *Average RR Signal Morphology*

The following steps adopted for the calculation of ARSM parameter.
Step1: Each RR divided into 300 equal parts and for each part; the signal amplitude is measured by applying spline interpolation method. This 300 dimensional data for each RR signal morphology is referred by us as RSM.
Step 2: The average of RSM's extracted from ECG signals gives the ARSM.

2.3 ***Classifier Selection***

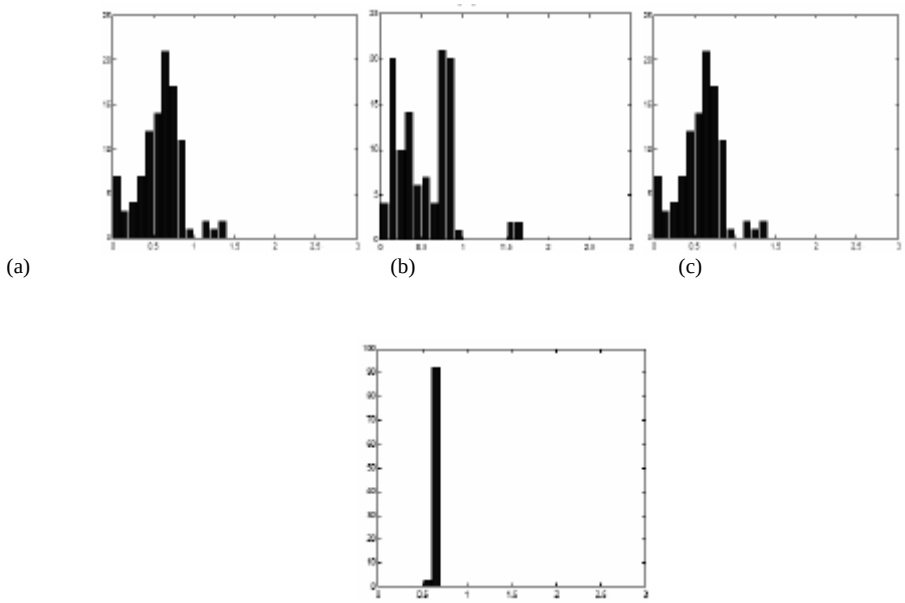
We have devised a new strategy to test the efficiency of feature parameter in classifying different group of data and we have applied it to check the efficiency of ARSM and RIFH to classify different disease group. For our study, we have chosen the disease classes as normal, paroxysmal arrhythmia and bradycardia. From the data of all the classes, we have calculated the corresponding ARSM and RIFH feature parameters. In the next step, a two-dimensional feature validation tool, employed to compare the efficiency of individual features. The tool works through the following algorithm:
Step 1: For each class the average of the feature of interest is calculated and considered as the CG (i.e. center of gravity) of that class.
Step2: Maximum distance of the feature point of this class is calculated and considered as radius of the class-sphere for the particular class of interest. We term them as class radius.
Step 3: First, a pair of classes is considered. Then CG-to-CG distance between this pair is calculated and plotted as the bar diagram for this pair. The two ends of the bar, considered as two CG" s of the two classes. Two circles of respective class-radius drawn considering these bar-ends as two circle centers.
Step 4: Step 3, followed for other pair of group also. In this study, for three pair of disease only three pairs are possible which, are shown as C1, C2, and C3.

3. Results and Discussions

In this current work, the methodology proposed, based on an automatic identification of RR interval and the study of signal pattern within this RR interval. RR interval of the ECG signal mostly used in the calculation of the heart rate of the patient and is one of the markers of one cycle of the ECG signal, which is generated from S-A(sino-atrial) node to S-A (sino-atrial) node [7-11]. The R peak of the ECG signal generated from the contraction of the ventricles and applied in different disease identification. The main reason for considering it for our work is that it gives higher amplitude value in the plot of ECG signal, and so easily identifiable. The R-to-R wave of an ECG signal provides not only the information of the Heart Rate from the ECG signal but the functional style of heart also.
In this study we are basically doing two level signal processing. One is to recognize the RR-peak and second one is to identify disease from the RR-signal morphology based feature parameters. In this study we are using two feature parameters. One is frequency histogram of RR interval and another is average signal morphology of the RR interval of highest frequency.

3.1. *RR interval Frequency Histogram (RIFH)*

The discriminatory power of a simple feature like RIFH is shown in the figure given below.

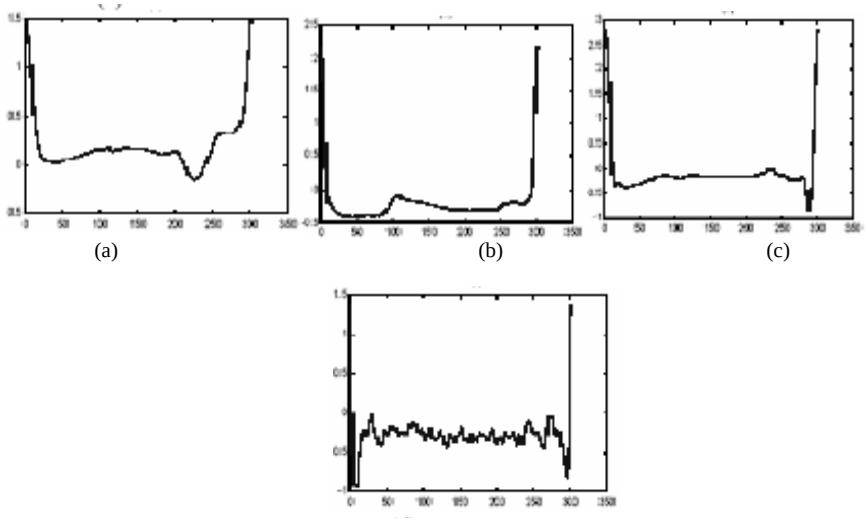


(a) (b) (c)

(d)

Fig. 3. Frequency Histogram of RR interval: (a) normal (b) CU Ventricular Tachycardia (c) Bradycardia and (d) paroxysmal arrhythmia

3.2. *Average RR Signal Morphology*



(a) (b) (c)

(d)

Fig. 4 Average RR Signal Morphology: (a) normal (b) bradycardia (c) Tachycardia (d) paroxysmal

3.3. *Feature validation tool*

The efficiency of the features ARSM and RIFH can be visually judged by looking into the Fig. 5 and 6 respectively. Fig. 6 clearly indicates that for building a classification scheme the feature RIFH should be more preferable. This is because we get all the class-pairs as non-overlapping for this feature, which indicates that these classes are fully separated in the multidimensional space.

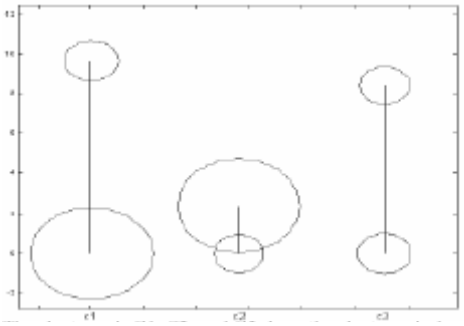


Fig. 5 The cluster pair C1, C2, and C3 shows the cluster pair drawn from Average RR signal Morphology of normal & paroxysmal arrhythmia, paroxysmal arrhythmia and bradycardia, bradycardia and normal.

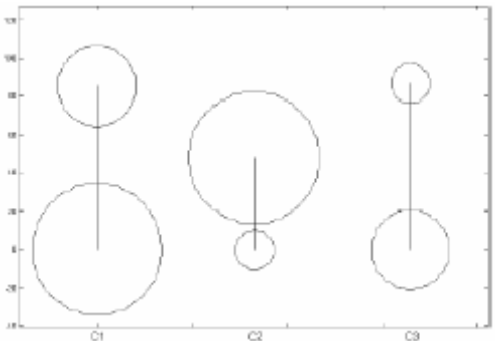


Fig. 6 The cluster pair C1, C2, and C3 shows the cluster pair drawn from RR interval Frequency

4. Conclusion

In this study our main objective is maximum exploitation of information from apparently not-so-informative and regular bio-rhythmic signals like ECG. This has been done to check whether information potentiality of a very regular type of signal can be enhanced through the design of the above described simple but novel data mining method. Another motivation for this type of work is to utilize the signal in a holistic sense which clinicians generally discard due to lack of known tool to tackle such whole data. In this direction we mined two new feature-parameters like RR Interval Frequency Histogram (RIFH) and Average RR Signal Morphology (ARSM). We found that these two features could significantly and very efficiently discriminate different cardiac arrhythmias. The straightforward feature validation tool designed by us appeared to be helpful in further confirming the above said efficiency of the newly designed features.

Acknowledgement

I am thankful to Dr. David J. Christini, Assistant Professor Cornell Univ., NY, USA for giving initial clue of data source.

References

[1] Prokhorov M.D., Ponomarenko V.I., Gridnev V.I., Bodrov M.B., Bespyatov A.B (2003). Synchronization between main rhythmic processes in the human cardiovascular system. Phys. Rev. E, V.68, 041913.

[2] Bezruchko, B.P., Yu.V.Gulyaev,S.P.Kuznetsov, and E.P.Seleznev(1986). New type of critical behavior of coupled systems at the transition to chaos. Sov.Phys.Dokl, 31(3): 258-260.

[3] Aparna M. Surve, Prem C. Pandey,S. S. Pratapwar, Vinod K. Pandey, N. S. Manigandan (2004). Enhancement of Radial Arterial Pulse by Spectral Subtraction. ICSCI Hyderabad, India.
[4] Lepage, Ronan, Boucher, Jean-Marc,Blanc, Jean-Jacques(2001). ECG Segmentation and P-Wave Feature
[5] Extraction Application to Patients Prone to Atrial Fibrillation. Conference paper, 25.
[6] Surehka Palreddy. ECG BEATS DATABASE Description, 1996. (book)
[7] H. Gholam-Hosini and H. Nazeran(1998). Detection and Extraction of ECG signal parameter. IEEE Engineering in Medicine and Biology Society, 20(1).
[8] Traditional Tibetan Medicine http://www.tararokpa.org/diagnose.html#pulse
[9] Representation of different heart blocks http://www.guidant.com/condition/heart/heart_chambers.shtml
[10] For different Abnormalities in ECG signal and representation of ECG Signal http://medlib.med.utah.edu/kw/ecg/mml/ecg_533.gif, http://medlib.med.utah.edu/kw/ecg/ecg_outline/Lesson4/index.html
[11] For database http://www.physionet.org