International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Automated Coronary Artery Disease Detection using RQA Features and Quadratic Support Vector Machine

Saurav Mandal, Nabanita Sinha

Abstract*: One of the major causes of death globally due to heart disease is the coronary artery disease (CAD). Due to CAD the blood flow to the cardiac muscle is reduced. The progression of this process eventually causes Myocardial Infarction (MI) that result in sudden death. Hence the detection of CAD at early phase is essential. The electrocardiogram (ECG) is mainly used to capture the abnormal cardiac activity for CAD. But the difficulties in manual interpretation of ECG signal leads to error in CAD detection. To overcome the difficulty in CAD diagnostic task, we have proposed a computer aided methodology using the heart rate variability signal (HRV) for auto diagnosis of CAD and Normal heart condition. The hidden characteristics of HRV signal are identified through Recurrence Plot (RP) and the hidden information is quantified by Recurrence quantification analysis (RQA). The extracted RQA based nonlinear features are analyzed for their clinical significance. The set the effective features are used for classification and subjected to three types of Support vector machine (SVM) classifier to discriminate CAD and the normal heart condition. The ECG database of CAD and normal subjects are taken from Physio.net database to obtain the experimental results. The highest diagnostic ability of the classifier is obtained by quadratic SVM with the accuracy of 98.83% where as the linear and cubic SVM classifier provide 97.22 % and 98.37 % classification accuracy respectively.

***Keywords* : Coronary Artery Disease, ECG, Recurrence Plot , Recurrence Quantification Analysis, Quadratic SVM**

I. INTRODUCTION

Coronary artery disease (CAD) is one of the primary causes of death worldwide, accounting for more than 17.6 million deaths per year and expected it will rise to more than 23.6 million by 2030, according to a 2014 study [1]. According to American heart association statistics report in 2019 suggest that 840678 deaths in US in year 2016, approximately 1 of every 3 deaths due to CAD [1]. Electrocardiogram (ECG) is normally used for CAD and it is a technique that records electrical action of heart over time. In clinical diagnosis several techniques is used to detect CAD from ECG signal but all those techniques have at least one limitation such as longer examination time, higher cost, low sensitivity and specificity. Due to low sensitivity all patients ECG signal not to reach target heart rate [3]. Non-invasive diagnostic techniques have been introduced to overcome those limitations of clinical diagnosis techniques. Non-invasive diagnosis techniques are capable to detect CAD in early stages. Non-invasive nature of HRV signal is the indication of some cardiac dysfunction. Time domain analysis value of HRV signal is significantly higher in normal patients compare to CAD patients [4]. Autoregressive power spectral density of HRV signal also correlated to angiographic severity of CAD patient [5]. Zia et.al [6] has shown in his research article the strategy of noise detection and exclusion for acoustic CAD identification. Various methods have widely studied on revealing of CAD from ECG signals [7-9] time and frequency analysis using different features extraction methods like Pointcare plot [12-14], Hurst exponents [15], discrete wavelet transform[16], linear discriminant analysis [17], and higher order spectra (HOS) [18]. Though many research works have been reported on CAD detection, but detecting effective features for decreasing classifiers response time is still a challenge.

In our proposed work we have estimated nonlinear features to analyse the hidden characteristics of CAD HRV signals. The extracted RQA parameters are subjected to three different classifiers for achieving high diagnosis accuracy of CAD. The support vector machine classifier is used to achieve better accuracy to detect CAD subjects from HRV signal.

II. METHOD

Our proposed methodology for automated CAD diagnosis is mainly consists of four stages. The flow chart of our proposed model is shown in Fig. 1. First stage is the pre-processing of the input raw ECG signal taken from the MIT database. The second stage deals with the segmentation of HRV signal. The next stage is extraction of significant features from HRV signal. In the last stage is the classification of CAD condition and normal sinus rhythm (NSR). The significant feature extractions are very important to increase detection accuracy of the classifier. Thus, we have extracted RQA based nonlinear features in this study. All features extraction methods are described below.

Manuscript published on January 30, 2020.
* Correspondence Author
Saurav Mandal*, PhD Research Scholar, Razabazar Science College, University of Calcutta, Kolkata, India. Email: souravsourav.cu@gmail.com
Nabanita Sinha, PhD Research Scholar, Razabazar Science College, University of Calcutta, Kolkata, India. Email:nabanitaroy.sinha@gmail.com

A. Data Acquisition

The open access database of physio.net has been used for the validation of our proposed work [6]. We have taken R-R interval of normal ECG data signal from MIT-BIH database and CAD data signal from PTB diagnostic database.
Each ECG is sampled at frequency of 128 Hz. The RR-interval database of CAD is used for our experiment. In this study, we have used 30 dataset of CAD patients and 30 dataset NSR subjects.

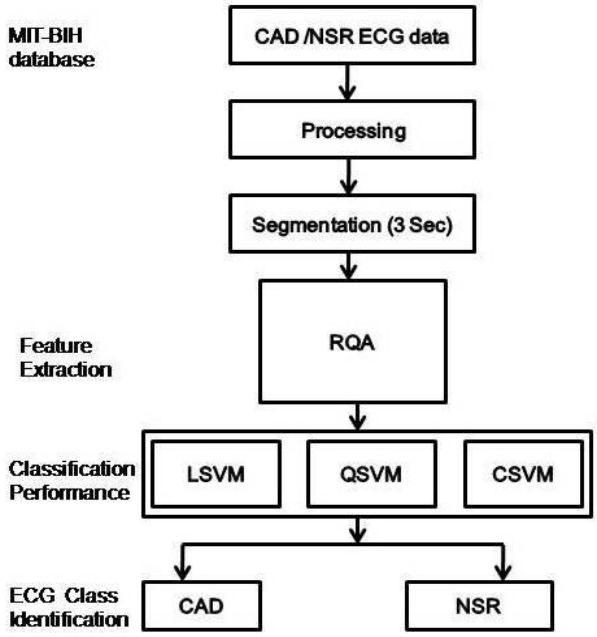

Fig. 1. Flow plan of designed classification system

B. Preprocessing

In this work preprocessing of input ECG signal is the noise removal stage. This stage is introduced before segmentation of input signal to increase the efficiency of the system .This stage removes both the baseline wandering and power line interference noise from HRV signal. The low frequency and high frequency component are eliminated by using filter having cut-off frequency of 0.4 Hz and notch filter of 43 Hz respectively [17].

C. Segmentation

After the pre-processing stage HRV signals are segmented for feature extraction and analysis. The HRV signals are taken of duration 1hr from physio.net database. The segmentation is done by taking input HRV signal where each patient's data set contains 462255 samples for CAD data and 493038 samples for NSR data. All records are sampled with cutoff frequency 43 Hz. One window in CAD data number of samples contains 43*4=172 and window size is 4sec. In NSR data number of samples is contain 43*2=86 and window size is taken 2sec. We have found total numbers of segment in CAD patients are 2674 and NSR patients are 5733.

D. Recurrence Quantification Analysis

Recurrence plot (RP) is the graphical representation of phase space path which is suitable for identification of hidden patterns and nonlinearities present in time series data. If the distance between the state p and state q is less than the predefined value then recurrence is occurred. Let y_p and y_q are two points in the trajectory of n -dimensional space. When distance between y_p and y_q is very close, then the recurrence id followed and location of spot is at position p*q. The plot is symmetric through the diagonal path. RP is visualized ad patch of black and gray dots of M*M dimension. The dark dot indicates the recurrence. The hidden information present in RP time series is quantified by Recurrence quantification analysis (RQA) which represents numeric quantities of recurrence plot. The RQA features are estimated to measure the complexity and non stationary dynamics of HRV signal [22]. The RQA parameters are explained below.

- Recurrence Rate (RR): Recurrence represents the percentage of the recurrence spots (M) or denoted as correlation sum. Recurrence rate is counted from the dark dot in RP.

$$RR = \frac{1}{M^2}\sum_{p,q=1}^{M} R_{p,q} \tag{1}$$

- Determinism (DET) : It is measure in based on diagonal lines. DET is measure in ratio of recurrence points in diagonal structure to all recurrence points.

$$\text{DET} = \frac{\sum_{l=d_{\min}}^{N} lD(l)}{\sum_{pq}^{N} R_{p,q}} \tag{2}$$

$$\text{Where } D(l) = \sum_{p,q=1}^{N}(1-R_{p-1,q-1}(1-R_{p+1,q+1}))\prod_{k=0}^{l=1} R_{p+k,q+k}$$

is the histogram of the intervals, l of the crosswise lines and d_{min} parameter sets the lower bounds on diagonal lines in the recurrence plot.

- Ratio : It is proportion of determinism and recurrence rate. It is computed based on crosswise line $D(l)$ of length l as follows:

$$\text{Ratio} = \text{N}^2 \frac{\sum_{l=d_{\min}}^{N} lD(l)}{(\sum_{l=1}^{N} D(l))^2} \tag{3}$$

Dynamic ratio denotes the amount of transition occurred in the system.

- Maximum Diagonal line length (L_{max}) : Diagonal lines gives an hint regarding the digression of the trajectory sectors. Small value of L_{max} indicates move deviation in the trajectories.

$L_{max} = \arg\max_L D(l)$ (4)

- Mean Diagonal Line length (L_{mean}): It is parameter which measures the mean length of diagonal lines in RP.

$$L_{mean} = \frac{\sum_{l=d_{min}}^{N} lD(l)}{\sum_{l=d_{min}}^{N} D(l)} \quad (5)$$

L_{mean} is represented the mean of predication time.

- Entropy (ENT) : It is a parameter which represent the complexity of dynamic system. System deterministic structure complexity is also increased respect to the higher value of entropy.

$$ENT = \frac{D(l)}{\sum_{l=d_{min}}^{N} D(l)} \quad (6)$$

- Maximum Vertical line length (V_{max}) : This parameter is measure the longest vertical line length in recurrence plot.

$$V_{max} = \arg\max_L D_v(l) \quad (7)$$

III. CLASSIFIER

A. Quadratic Support Vector Machine

SVM is supervised learning method that analysed data for classification and regression. Two classes are separated using decision boundary and it is called hyperplanes. We consider n training data pairs P_i, Q_i. Where P_i input vector and Q_i is target vector of input data.

$$P_i = \begin{bmatrix} P_{11}P_{12}P_{13}.....P_{1d} \\ P_{21}P_{22}P_{23}.....P_{2d} \\ P_{31}P_{32}P_{33}.....P_{3d} \\ ..... \\ P_{n1}P_{n2}P_{n3}.....P_{nd} \end{bmatrix} \quad Q_i = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ .... \\ Q_d \end{bmatrix} \quad (8)$$

Hyperplane (S,C) is capable of separating two data point in two classes is given :

$S^T P + C = 0$ (9)

where $S=[s_1, s_2, s_3, s_{4,.....}, s_d,]^T$ and C is scalar

Class 1 is labelled by +1 and -1 is represented class 2.

$Q_i = +1$ if $S^T P + C \geq 0$ and $Q_i = -1$ if $S^T P + C < 0$,

$Q_i(S^T P + C) \geq 1$, (10)

$Q_i(S^T P + C) = \gamma^*(i) \geq 1$ (11)

$\gamma^*(i)$ is called the functional margin. Good classification is considered if $\gamma^*(i) \geq 1$.

$$\text{Geometrical margin } \gamma(i) = \frac{\gamma^*(i)}{\|S\|} \quad (12)$$

SVM optimization problem is occurred to maximize geometrical margin in subject to functional margin greater than $\gamma^*(i)$. Linear separation of large data is difficult using linear SVM in high dimensional feature space.

Quadratic programming is the process to solve linearly constrained quadratic optimization problem. Quadratic SVM is capable of separating nonlinear data in two or more classes are as follows:

$$f(P) = \frac{1}{2} P^T W P + C^T P + E \quad (13)$$

where quadratic function W,C,E

$$W = \begin{bmatrix} w_{11}w_{12}w_{13}.....w_{1d} \\ w_{21}w_{22}w_{23}.....w_{2d} \\ w_{31}w_{32}w_{33}.....w_{3d} \\ ..... \\ w_{n1}w_{n2}w_{n3}.....w_{nd} \end{bmatrix} \text{ and } C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ .... \\ c_d \end{bmatrix} \quad (14)$$

The nonlinear term

$$f_{nonlinear}(P) = \frac{1}{2} P^T W P \text{ and } f_{linear}(P) = C^T P + E \quad (15)$$

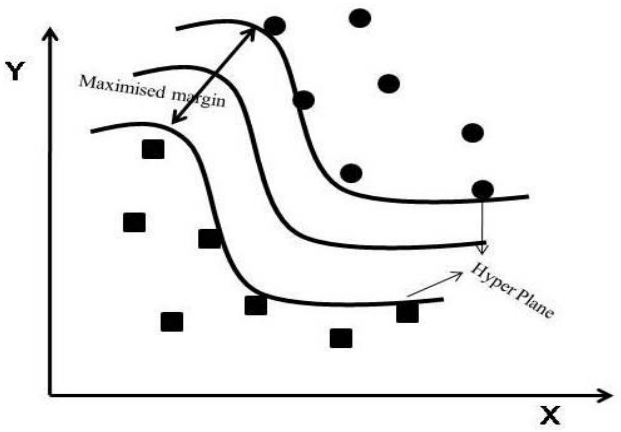

Fig. 2.Maximum margin of two hyperplanes for QSVM and trained samples of two classes

To separate the classes in best possible way cubic SVM is find a hyperplane in multidimensional feature space. Karnel function of cubic SVM classifier is $K(P_i, P_j)$ and

$K(P_i, P_j) = (P_i^T, P_j)^3$ (16)

For the duality property of QSVM classifier, it is more capable to detect coronary artery disease

IV. RESULT

The ECG signals of two different databases are pre-processed with total 462255 samples in CAD and 493038 samples in NSR data. All segments of CAD and NSR are transformed independently to the higher dimensional RP plots. The plots depict the dynamic pattern of the time series data. Recurrence plots in Fig 3-4 represent the natural time correlation in the ECG signals. RP denotes the degree of similarity present in the given signal and it is almost symmetric with diagonal. We have extracted seven RQA features using these recurrence plots. Table I shows the obtained experimental result.

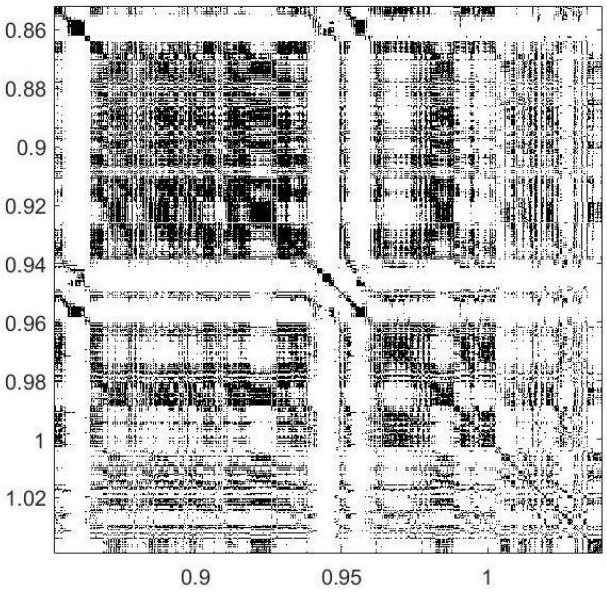

Fig. 3.Recurrence Plot (RP) of NSR data

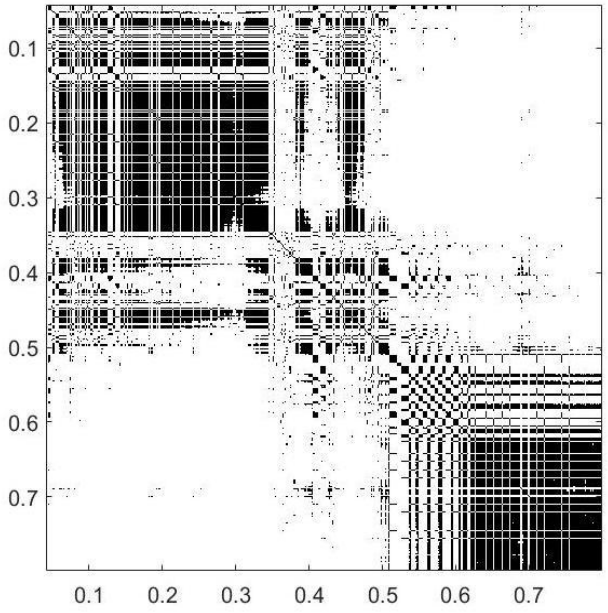

Fig. 4.Recurrence Plot (RP) of CAD data.

We have fed the RQA parameters to the SVM classifier. These studies, three different types of SVM classifier are used. The performance of linear SVM, quadratic SVM and Cubic SVM are analysed independently. The performance of individual classifier is estimated by calculating the percentage of accuracy, sensitivity and specificity. Fig 5-7 represent the performance analysis of three types of SVM classier used in our proposed classification model.

Table- I: Nonlinear features range for the normal and CAD data

Data Types	Extracted Features						
	RR	***DET %***	***Ratio***	L_{MAX}	L_{mean}	***Entropy***	V_{max}
NSR	504.47 ± 20	99.70 ± 05	0.0030 ± 0.0010	363.83 ± 20	68.10 ± 10.50	6.48 ± 0.20	36.23 ± 10
CAD	347.73 ± 30	99.50 ± 10	0.0040 ± 0.0010	210.78 ± 30	52.50 ± 12.50	5.98 ± 0.12	38.53 ± 10

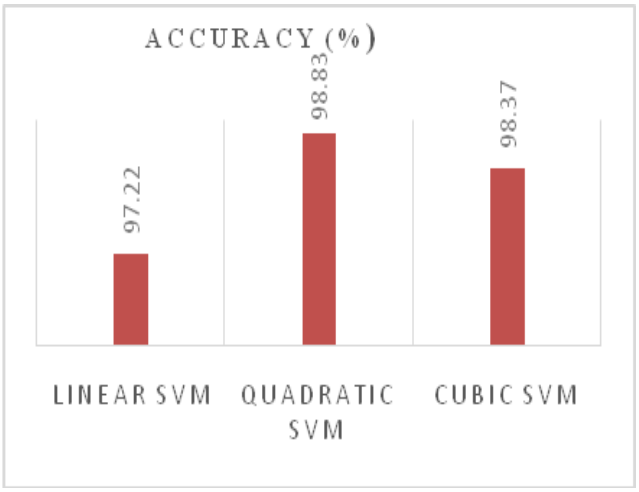

Fig. 5.Accuracy of proposed classification model

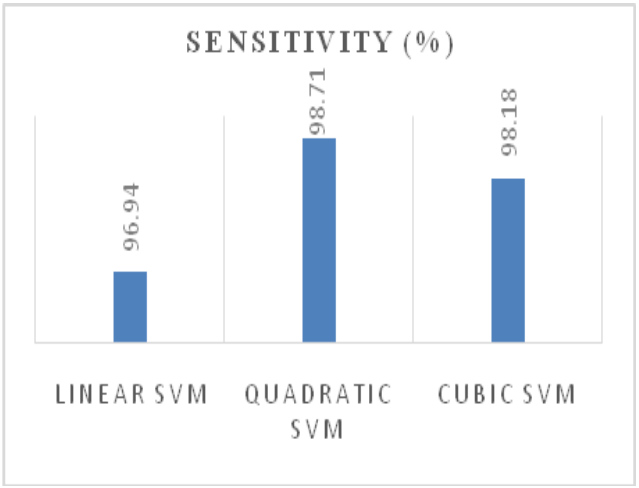

Fig. 6.Sensitivity of proposed classification model

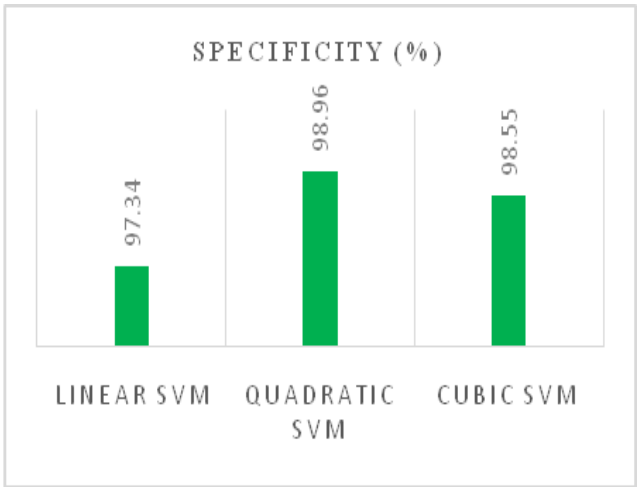

Fig. 7.Specificity of proposed classification model.

V. DISCUSSION

In this study, the RQA based nonlinear features are analyzed to capture the different changes occur in HRV signal for CAD and normal condition of heart. The recurrence plots as shown in Fig 3-4 explicitly represent distinguishable characteristics for two types of ECG signals. The extracted features are able to pick up the sudden variation occurs in cardiac activities during CAD. It is observed from Table I , the obtained seven features provide distinctive values to separate CAD and normal classes .The extracted seven potential features are fed the three different types of SVM classifiers to automatically detect CAD. The most efficient classification result is obtained from the Quadratic SVM classifier. The maximum accuracy of 98.83 % sensitivity value of 98.71% and specificity of 98.96% are obtained from quadratic SVM classifier. The linear SVM provides the minimum accuracy of 97.22 % and 98.37 % classification accuracy is obtained by the cubic SVM classifier. We have also compared our work with the various studies based on CAD diagnosis. The Table II provides the comparative summary of various research works.

Table- II: Comparison of proposed work with previous studies

No.	Author	Classifier	Accuracy(%)
1	Lee et.all [20]	SVM	90
2	Dua et. all [22]	GMM	89.8
3	Giri et. all [21]	MPL	96.8
4	Acharayaet.all [19]	CNN	95.11
5	Proposed	Quadratic SVM	98.83

VI. CONCLUSION

Computer aided automated prediction of CAD is important to prevent its progression which may reduce heart attack. In this work we have explored the Recurrence plot for the analysis of HRV signal to diagnose CAD. The extraction of nonlinear RQA features is effective for the non-stationary nature of ECG signal. The RQA parameters provide hidden information about the HRV dynamics for CAD which differs from normal condition of heart. These extracted clinically significance features are utilized by SVM classifier. The performances of different SVM classifiers are estimated to achieve high accuracy for CAD detection. The maximum efficiency is obtained by the quadratic SVM classifier to separate two classes of ECG signals. Our proposed methodology has faster response time in detection of CAD for using less number of effective features.

REFERENCES

1. AS, Mozaffarian D, Roger VL, Benjamin EJ, On behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. Circulation.2013
2. Silber EN, Katz LN, Heart disease, MacMillan Publishing Co. Inc, NewYork, 1975.
3. S Dua, Xian Du, V Sree, T Ahamed, Novel classification of coronary artery disease using heart variability analysis, Journal of Mechanics in Medicine and Biology, vol-12, No-4, 1240017.
4. Bigger JT Jr., Fleiss JL, Steinman RC, Rolnitzky LM, Schneider WJ, Stein PK, RR variability in healthy, middle-aged persons compared with patients with chronic coronary heart disease or recent acute myocardial infaction, Circulation **91**: 1936-1943, 1995.
5. J Hayano, Y Sakakibara, A Yamada, N Ohte, T Fujinami, K Yokoyama, Y Watanabe, K Takata, Decreased magnitude of heart rate spectral components in coronary artery disease: Its relation to angiographic severity, Circulation **81**:1217-1224, 1990.
6. Zia MK, Griffel B, Semmlow J, Fridman V, Saponieri C, Noise detection and elimination for improved acoustic detection of coronary artery disease, Journal of Mechanics in Medicine and Biology **12**(2)1240007,2012.
7. Acharya UR, Sudarshan VK, Koh Jew, Martis RJ, Tan JH, Oh SL, Adam M, Hagiwara Y, Mookiah MRK, Chua KP, Chua KC,Tan RS, Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals, Biomedical Signal Processing Cont 31:31-43, 2017.
8. Arafat S, Dohrmann M, Skubic M, Classification of coronary artery disease stress ECGs using uncertainty modelling, IEEE Cong Comput Intell Meth Appli, Istanbul,Turkey,2005.
9. Babaoglu I, Findik O, Bayrak M, Effects of principle component analysis on assessment of coronary artery diseases using support vector machine, EXP Syst Appl 37:2182-2185,2010.
10. Kumar M, Pachori RB, Acharya UR, Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals, Biomedical Signal Process Control 31: 301-308,2017.
11. Schreck DM, Ng L, Schreck BS, Bosco SF, Allegra JR, Zacharias D, Wortzel JV, Detection of coronary artery disease from the normal resting ECG using nonlinear mathematical transformation, Ann Emerg Med 17:132-134, 1988.
12. Kim WS, Jin SH, Park Yk, Choi HM, A study on development PF multi-parameter measure of heart rate variability diagnosing cardiovascular disease, Int Fed Med Biol Engg Proc 14: 3480-3483, 2007.
13. Lee Hg, Noh KY, Ryu KH, A data mining approach for coronary heart disease prediction using HRV features and carotid artial wall thickness, IEEE Int Conf Biomed Eng Inform, Sanya, China, 2008.
14. Lee Hg, Noh KY, Ryu KH, Mining biosignal data, Coronary artery disease diagnosis using linear and nonlinear features of HRV, Springer-Verlag, Berlin Heidelberg, pp.218-228, 2007.
15. Zbilut JP, Webber Jr. CL, Embedding and delays as derived from quantification of recurrence plots, Phys Lett A 171:199-203, 1992.
16. Patidar S, Pachori RB, Acharya UR, Automated diagnosis of coronary artery diasease using Tunable-Q wavelet transform applied on heart rate signals, Knowl-Based Syst 82: 1-10, 2015.
17. Wariar R, Eswaran C, Integer coefficient bandpass filter for the simultaneous removal of baseline wander, 50 and 100 Hz interference from the ECG, Med Biol Engg Comput 29(3): 333-336,1991.

18. Huikuri HV, Makikallio TH, Peng CK, Goldberger AL, Hintze U, Moller M, Fractal correlation properties of R-R interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction, Circulation **101**:47-53,2000.
19. Acharya UR, Fujita H, Oh SL, Muhammad A, Tan JH, Chua KC, Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network, Knowl-Based Syst 132:62–71, 2017.
20. H.E.Lee ,K.Y.Noh, K.H.Ruy, Miningbiosignal data: coronary artery disease prediction using HRV features and carotid arterial wall thickness, in : international conference on Biomedical Engineering and Informatics, vol 1,pp.200-206,2008.
21. D.G.Giri, U.R.Acharya, R.J,Martis, V.C.Sree, Automated diagnosis of coronary artery disease affected patients using LDA, PCA,ICA and discrete wavelet transform, Knowledge based system ,2012.
22. S.Dua, X.Du, V.S.Sree. T.V.I.Ahmed, Novel classification of coronary artery disease using heart rate variability analysis, Journal of Mechanics in Medicine in Biology 12(4) 2012, 1240017-1-19

AUTHORS PROFILE

Asst. Prof. Saurav Mandal (HOD-ECE) from Maulana Abul Kalam Azad University of Technology, completed Bachalor of Technology in Electronics and Communication Engineering and Master of Technology from Maulana Abul Kalam Azad University of Technology (In-house). Currently pursuing PhD from Razabazar Science College, University of Calcutta. His area of interest are Biomedical Signal Processing, Nonlinear Dynamics, Machine Learning, Pattern Recognition and VLSI design.

Asst. Prof. Nabanita Sinha from Maulana Abul Kalam Azad University of Technology, completed Bachalor of Engineering in Electronics and Communication Engineering from University of Burdwan. Completed Master of Technology from University of Calcutta. Currently pursuing PhD from Razabazar Science College, University of Calcutta. Her research areas are Biomedical Signal Processing, Nonlinear Dynamics, Machine Learning and Pattern Recognition.