

Classification of Heart Arrhythmia in ECG Signals using PCA and SVM



Sumanta Kuila, Namrata Dhanda, Subhankar Joardar

Abstract: Electro cardiogram (ECG) signals records the vital information about the condition of heart of an individual. In this paper, we are aiming at preparing a model for classification of different types of heart arrhythmia. The MIT-BIH public database for heart arrhythmia has been used in the case of study. There are basically thirteen types of heart arrhythmia. The Principal Component Analysis (PCA) algorithm has been used to collect various important features of heart beats from an ECG signal. Then these features are trained and tested under Support Vector Machine (SVM) algorithm to classify the thirteen classes of heart arrhythmia. In the paper the proposed algorithm has been discussed and the outcome results have been validated. The result shows that the accuracy of our classifier in our research work is more than 91% in most of the cases.

Keywords: Arrhythmia, Electrocardiogram, MIT-BIH database, Principal Component Analysis, Support Vector Machine.

I. INTRODUCTION

An ECG is a complete representation of the electrical activity of the heart on the surface of the human body, and it is extensively applied in the clinical diagnosis of heart disease [1]. Because of problems in the heart, many people die every year. The ECG is a graphic record of the direction and magnitude of the electrical activity that is generated by depolarization and repolarization of the atria and ventricles. Figure -1 describes a standard diagram of an ECG signal, its different components, different peaks with the distances among the peaks. The most important part is QRS complex which will give support to analyze the features of the signal and the relevant classification can be done from there. ECG is a non-invasive technique and is recorded by applying the device to a patient when they need to monitor patients' ECG to find the few abnormal cycles. One cycle in an ECG signal consists of P-QRS-T parts. Most of the useful information in the ECG is found in the intervals and amplitudes defined by its characteristic wave peaks and time durations. Here ECG feature extraction is the most important thing and developing quick and accurate methods for automatic ECG feature extraction. Various studies have been done for the

classification of various arrhythmias [2][3]. One of the most important problem is to precisely detect the beats and classify them as normal and abnormal signals. Simply by looking at an ECG plot, it can be noticed that QRS complex is the predominant feature. Other features like P and T wave are sometimes too small which cannot be detected. Hence the main feature i.e. QRS complex will yield the best accuracy in detection of waves[4][5].

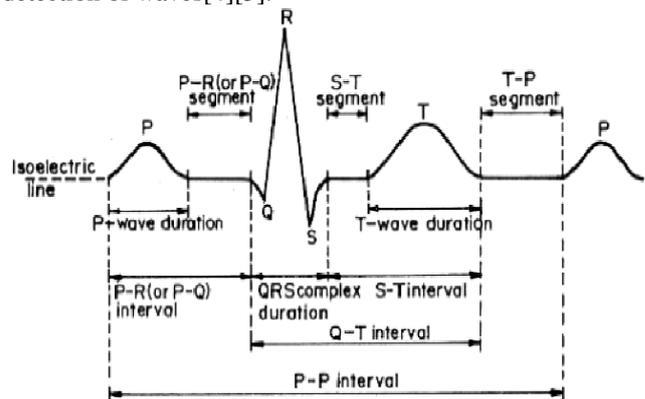


Fig -1: Standard ECG signal block diagram

The main objective of this paper is to propose automated computer-based system for beats classification using Principal Component Analysis (PCA) for feature extraction which reduces the dimensionality of signal and then use SVM (Support Vector Machine) to classify the ECG beats. To train the SVM and validate the result we have used the MIT-BIH Arrhythmia Database [6]. In this paper section II described the principal component analysis and its impact on signal classification, section III describes support vector machine, section IV describes support & methodology and the last section experiments and results.

II. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis is mainly used to reduce the dimensionality of the dataset from million values to specific values. So using PCA we can make sure to use the dataset easily, the algorithm removes noise from the dataset, it also reduces computational cost and at the understanding of the results become easier by this technique.

2.1 Basics of PCA

The basic goal of PCA is to reduce complexity of dataset for feature analysis where most important features are first identified. The main working principal is that it works by revolving the axes to align with the major variance in the dataset. The additional axes are selected orthogonal to the 1st axis, towards the direction of biggest variance. The analysis of Eigen value upon covariance matrix generates different set of orthogonal axes[7][8].

Revised Manuscript Received on August 30, 2020.

* Correspondence Author

Sumanta Kuila*, Department of Computer Science & Engineering , Haldia Institute of Technology, Haldia, West Bengal, India. E-mail: smanta.kuila@gmail.com

Namrata Dhanda, Department of Computer Science & Engineering ,Amity School of Engineering & Technology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, Uttar Pradesh, India. E-mail: ndhanda510@gmail.com.

Subhankar Joardar, Department of Computer Science & Engineering , Haldia Institute of Technology, Haldia, West Bengal, India. E-mail: subhankarranchi@yahoo.co.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



2.2 Impact of PCA on Feature analysis

Principal Component Analysis is a method which is statistical in nature that can reduce the dimension of the ECG signal (data) and is used to extract various types of features of an ECG signals. There are several techniques that can be used to identify hear arrhythmia but PCA is used as it can work precisely and effectively on the dataset. Moreover it can reduce the noise easily. PCA with respect to other give more importance to the variance structure of the new variables $Y_1, Y_2, Y_3, Y_4, \dots, Y_p$.

These variables are determined using different ranges so that the outcome becomes equal. Let the sample correlation matrix be R that is calculated from N observations on each of the principal component P of the above random variables. Let the pair of eigen-vector and eigen-values[9] of R be represented as

$$(X_1, E_1), (X_2, E_2), (X_3, E_3), (X_4, E_4), (X_5, E_5) \dots (X_p, E_p).$$

The i-th sample of the principal component of the given vector variable $Y = Y_1, Y_2, Y_3, Y_4, \dots, Y_p$ is given in the following equation---

$$E_i^W = E_{i1} W_1 + E_{i2} W_2 + E_{i3} W_3 + \dots + E_{ip} W_p$$

Where, $i = 1, 2, 3, \dots, p$

The i-th eigen-value is represented as:

$$E_i W = E_{i1}, E_{i2}, E_{i3}, E_{i4}, \dots, E_{ip}$$

The total sample variance in the principal component is equal to the all standardized variables. The standardized vector mathematical equation[10] is represented as:

$$W_v = (Y_v - \bar{Y}_v) / \sqrt{M_{vv}}$$

Where, $Y_v =$ sample mean . $M_{vv} =$ sample variance of variable Y_v

III. SUPPORT VECTOR MACHINE (SVM):

SVM is a strong discriminative classifier, which is formally described in a disconnect hyper plane. For supervised learning labeled training data is used in the shape of a optimized hyper plane is used to face both regression or classification challenges[11].

3.1 SVM method for Arrhythmia

The SVM classifier automatically select a proper subset of features to optimize the result. The below figure (Fig -2) Describes the generic SVM approach to classify the signal[12]. The figure (Figure 2) describes working procedure of the generic SVM approach. This describes the procedures based on SVM for feature extraction and classification. The procedure ensures quality classification and accurate feature extraction of the ECG signal. The noisy and negatively affected outcome would be avoided by using this procedure[13].

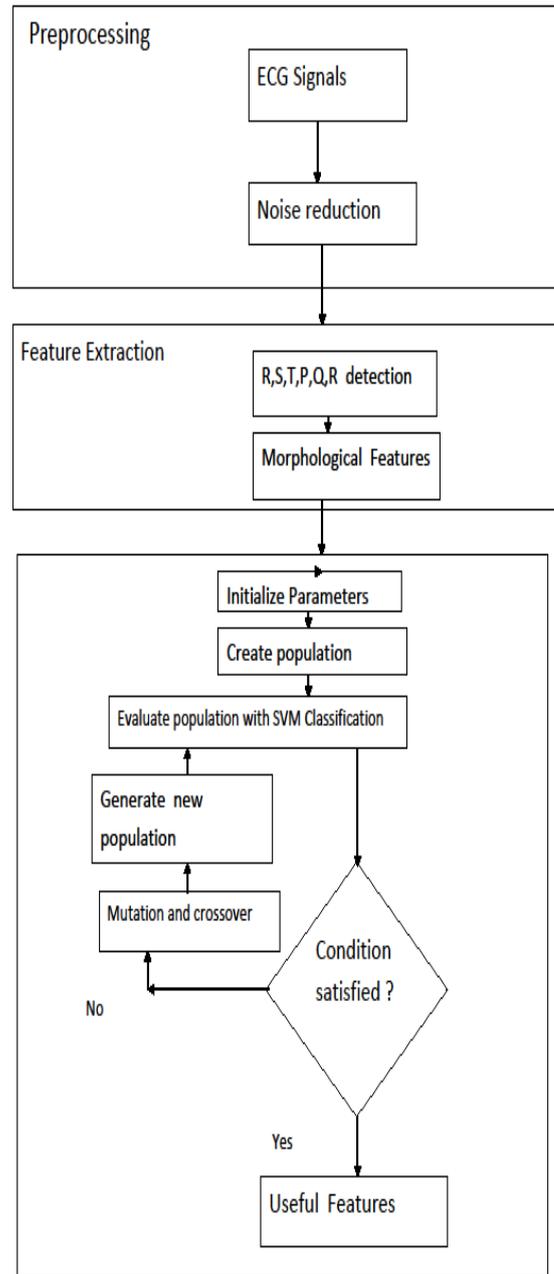


Figure 2 – The Generic SVM approach

3.2 Impact of SVM on classification

Support Vector Machines are used mainly for binary classification[14]. This can be easily be explained by mathematical expressions as follows below.

Let, the following equation be training set--
 $S = \{(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), \dots, (X_i, Y_i)\}$

Where,

$X_i =$ dimensional attribute

$Y_i = \{-1, +1\}$

i.e. $Y_i = -1$ and $Y_i = +1$ that are the two classes , class1 and class 2 which are being classified by the SVM.

Basically, SVMs can be denoted by a normal expression—
 $S(x) = P^T \phi(x) + b = 0$



Where,

$\Phi(x)$ is a mapping function.

P^T is a vector in feature set.

b is a scalar unit.

Here we are mainly using multiple class classifier behavior of the SVM[15][6]. Hence the typical expression for SVM including kernel function is given below:

$$g(x) = \text{sgn} \left\{ \sum_{i=1}^l a_i y_i \cdot K(x_i, x) + b \right\}$$

After extracting the features from the ECG signal of the MIT-BIH arrhythmia database the SVM classifier is used to classify the different signal. As SVM is used for binary classification for multiclass problem there are three different approaches that can be used[15]. They are as follows:

- One-against-one method
- One-against-all method
- Fuzzy decision making method

Here in our research we have trained the data set using cross validation and have used one against one method to find the result and accuracy of the 13 different types of heart arrhythmia.

The one-against-one method [16] can be depicted by the following expression:

$$D_i(x) = \sum_{j=1, j \neq i}^k \text{sgn}(D_{ij}(x))$$

IV. SETUP AND METHODOLOGY

The database used for classification is MIT-BIH arrhythmia database. Initially the records of the database are taken from the physionet. The collected test and train data are preprocessed and the unwanted noises are reduced with in certain level. Then using principal Component Analysis method the R – Peak are detected. The SVM classifier is used to classify it for feature extraction. The Figure -3 describes the whole procedure. The 13 different types of classified Arrhythmias are identified through out this process.

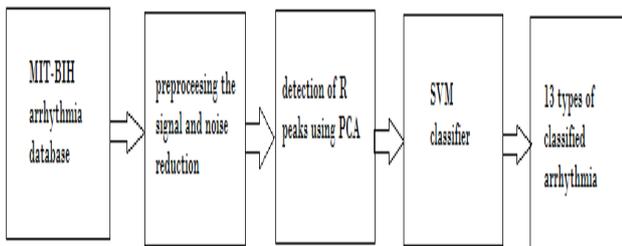


Figure- 3: Procedure of classification

Then, the segmented R peaks are passed under SVM classifier and we have detected 13 different types of heart arrhythmia using widely used classification one-against-one technique[17] and accuracy of our train and test sets have been examined. The outcome of our result is good as most of them resulted in more than 80% accuracy which implies that the classifier can be used in case of small data set also.

V. EXPERIMENT AND RESULTS

In the experiment, PYTHON v3.0 is used on PC with 2.00GHz in i3-6006U processor to run the desired model classifier and bring out the efficiency of our proposed model. In the experiment the MIT-BIH arrhythmia database[18] is used. The ECG signal of the database are firstly preprocessed and segmented into single heart beats. Then the processed signal is passed under PCA to determine the R peaks from the heart beats. In this experiment we have used the MIT-BIH arrhythmia database which is approximately consists of 48 data and 110,000 beats. The ECG signal of MIT-BIH is as plotted by our work is : The MIT-BIH database ECG signal that has been graphed by WFBD is also given below:

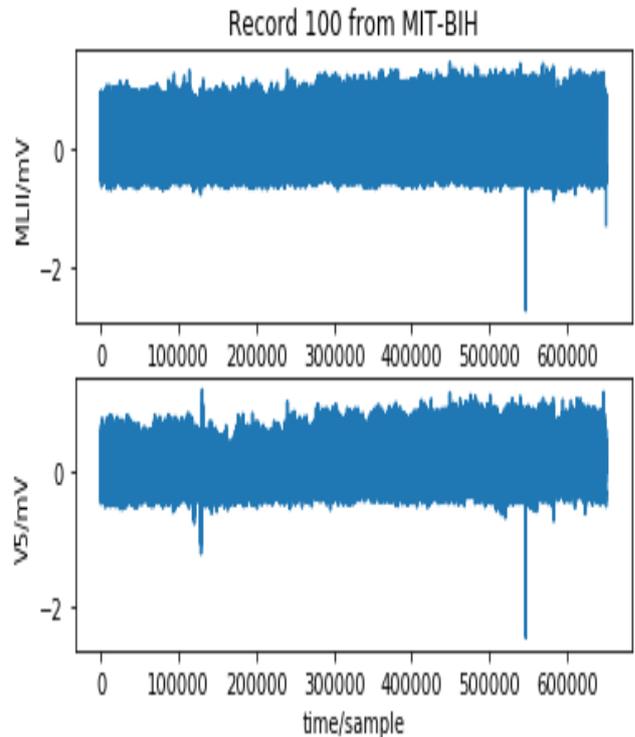


Figure -4: Plot of beat with time/sample

Figure 4 displays the outcome of the graph for beat outcome and the plot of beat with time sample of the records taken at the time of experiment. Through the graph plot the proposed SVM classifier is produced. The linear kernel performance is plotted and properly displayed. It shows the randomly generated general population upto time sample 600000 is produced. The diagram produces the subset of the features and it classify ECG signals with trained support vector machine. Those beats are generally classified into 13 different types of beats namely.

- Normal beat (N)
- Left bundle branch block beat (L)
- Nodal (junctional) premature beat (J)
- Right bundle branch block beat (R)
- Fusion of paced and normal beat (f)
- Premature ventricular contraction (V)
- Aberrated atrial premature beat (a)



Classification of Heart Arrhythmia in ECG Signals using PCA and SVM

- Atrial premature beat (A)
- Fusion of ventricular and normal beat (F)
- Nodal (junctional) escape beat (j)
- Ventricular flutter wave (!)
- Ventricular escape beat (E)
- Paced beat (\) [19][20]

The numbers of different types of beats present in the dataset are as follows. The MIT-BIH has total of 48 records and after compilation of whole data (both Test & Train) the different types of beats are generated.

- N = 10001
- L = 8075
- R = 7259
- e = 16
- j = 229
- A = 2546
- a = 150
- J = 83
- S = 2
- V = 7130
- E = 106
- F = 803
- / = 7028
- f = 982

Here it is normally classified each of the ECG signal into single heart beat segment. Then we centred each of the R spike from the beat signal and uses both the leads of the dataset to extract feature using PCA. Figure- 6 shows PQRS pick detection on record no 100 of train data and like this other picks are also detected and the training points are achieved.

The basic algorithm that has been used in the précised model have been depicted below:

1. Load the data and annotations
2. For Using PCA on the data, first scale the data this is done using StandardScaler function of SKLEARN.
3. Once the Scaling is done apply PCA which normalize the features by subtracting the mean and scale it to unit variance, using StandardScaler[21].
4. Then the process start of by selecting 3 orthogonal components.
5. After PCA is applied, split it data in train and test set.
6. Then put or data in SVM model to generate our classifier. SVM's C_value=1 generates the best result.
7. Then we use our test data to test our classifier[22].

The confusion matrix is then plotted. Now the description of R-peak of WFBD library of the python model helps the procedure to generate the confusion matrix. The whole data set is executed and the relevant outcome is generated here. .Figure 5 describes the confusion matrix as given the classified model.

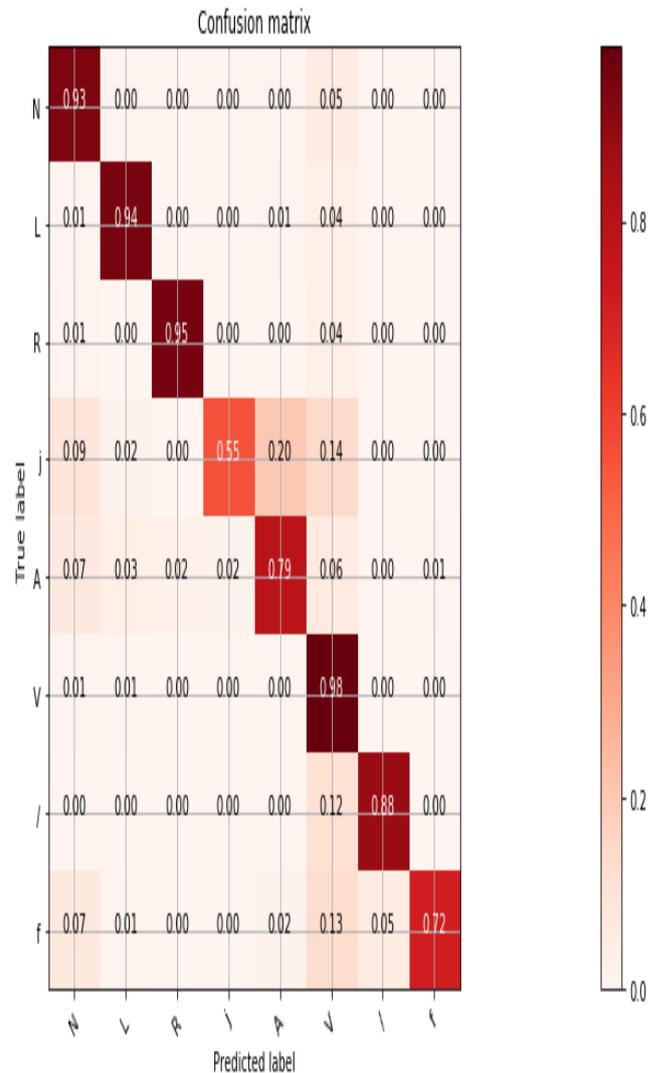


Figure-5 : The confusion matrix

Now the above diagram (diagram 5) plotted the confusion matrix of the model which shows a test accuracy of the model. The experiments shows the test accuracy is **92.142707%**. The table (Table-1) describes the simulation values of Precision, Recall, f1-score values of the metrics.

Table 1 : Precision , Classification report of sklearn metrics (SVM)

No	Precision	Recall	f1-score	Support
0	0.95	0.93	0.94	3283
1	0.97	0.94	0.96	2662
2	0.98	0.95	0.97	2460
3	0.68	0.55	0.61	65
4	0.90	0.79	0.84	889
5	0.76	0.98	0.85	2285
6	0.99	0.88	0.93	2293
7	0.92	0.72	0.80	330

Here the table shows the precision and classification report from sklearn matrices from support vector machine. It measures the precision, recall and f1-score and from there it derives the support.



Table 2 measures the accuracy, Macro Average and Weighted Average of the classification matrices. The table (Table-2) describes the description metrics of Accuracy, Macro Average, Weighted Average of the ECG dataset used.

Table 2: Accuracy, Macro average & Weighted average

	Precision	Recall	f1-score	Support
Accuracy	0.91	0.93	0.92	14267
Macro Average	0.89	0.84	0.86	14267
Weighted Average	0.93	0.92	0.92	14267

VI. CONCLUSION

In our research work we have used a strong and popularly used classifier i.e. SVM to classify various types of heart arrhythmia. We have used PCA for extraction of various features of an ECG signal and used a widely used classification technique i.e. one-against-one method to classify the extracted signals into different types of heart arrhythmia. In the result we have seen the accuracy in most of the cases is more than 80% which shows that our model can work well in case of small set of training data sets. But in some cases the accuracy have gone below 70% to 60% in our work. Hence in future we would try to improve our algorithm so as to create a model with high accuracy and also will be using other classification methods and compare which will be the best classification technique for heart arrhythmia detection.

REFERENCES

1. Syed Muhammad Anwar, Maheen Gul, Muhammad Majid, Majid Alnowami, "Arrhythmia Classification of ECG Signals Using Hybrid Features" Hindawi, Computational and Mathematical Methods in Medicine Volume 2018, Article ID 1380348.
2. Narendra Kohli, Nishchal K. Verma, and Abhishek Roy, "SVM based Methods for Arrhythmia Classification in ECG" International conference on Computer & Communication Technology, pp.486-490, IEEE 2010.
3. Arjon Turnip, M. Ilham Rizqywan, Dwi E. Arjon Turnip, M. Ilham Rizqywan, Dwi E. Kusumandari, Mardi Turnip, Poltak Sihombing, "Classification of ECG signal with Support Vector Machine Method for Arrhythmia Detection", IOP Conf. Series: Journal of Physics: Conf. Series 970 (2018), ICIESC-2017.
4. Raghu Nanjundegowda, Vaibhav Meshram, "Arrhythmia recognition and classification using kernel ICA and higher order spectra", International Journal of Engineering & Technology, Vol 7, pp.256-262, 2018.
5. Jalal A. Nasiri, Mahmoud Naghibzadeh, H. Sadoghi Yazdi, Bahram Naghibzadeh, "ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm", Third UKSim European Symposium on Computer Modeling and Simulation, pp. 187-192, 2009.
6. Mohammad Sarfraz, Ateeq Ahmed Khan, Francis F. Li, "Using Independent Component Analysis to Obtain Feature Space for Reliable ECG Arrhythmia Classification", IEEE International Conference on Bioinformatics and Biomedicine, pp. 62-67, 2014.
7. Yang Wu, Liqing Zhang, "ECG Classification Using ICA Features and Support Vector Machines", Springer-Verlag Berlin Heidelberg, ICONIP 2011, Part I, LNCS 7062, pp. 146-154, 2011.
8. Maedeh Kiani Sarkaleh, Asadollah Shahbahrani, "Classification of ECG Arrhythmia using Discrete Wavelet Transform and Neural Network", International Journal of Computer Science, Engineering and Applications, Vol.2, No.1, pp 1-13, February 2012.
9. Siva A, Hari Sundar M, Siddharth S, Nithin, Rajesh CB, "Classification of Arrhythmia using Wavelet Transform and Neural Network Model", Journal of Bioengineering & Biomedical Science, Volume 8, Issue 1, April 2018.

10. Parham Ghorbanian, Ali Ghaffari, Ali Jalali1, C Nataraj, "Heart Arrhythmia Detection Using Continuous Wavelet Transform and Principal Component Analysis with Neural Network Classifier", Computing in Cardiology, Vol. 37, pp.669-672, 2010.
11. Yinsheng Ji, Sen Zhang, Wendong Xiao, "Electrocardiogram Classification Based on Faster Regions with Convolutional Neural Network", Sensors, MDPI, Sensors 2019.
12. Vijaya Arjunan R, "ECG Signal Classification based on Statistical Features with SVM Classification", International Journal for Advance Signal & Image, Vol. 2, No.1, pp. 5-10, 2016.
13. Harpreet Kaur, Rupinder Kaur, "ECG Arrhythmia Detection using PCA and Elman Neural Network", International Journal of Science and Research, Volume 3 Issue 10, pp. 894-896, October 2014.
14. Qibin Zhao, Liqing Zhang, "ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines", IEEE, pp.1089-1092, 2005.
15. Usha Desai, Roshan Joy Martis, C. Gurudas Nayak, K. Sarika, Sagar G. Nayak, "Discrete Cosine Transform Features in Automated Classification of Cardiac Arrhythmia Beats", Emerging Research in Computing, Information, Communication and Applications, Springer, pp.153-162, 2015.
16. Yasushi Kikawa, Koji Oguri, "A Study for Excluding Incorrect Detections of Holter ECG Data Using SVM", Springer-Verlag Berlin Heidelberg 2004, ICONIP 2004, LNCS 3316, pp. 1223-1228, 2004.
17. Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, "Complex Physiologic Signals Physio Bank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231, May 2014.
18. Hongqiang Li, Danyang Yuan, Xiangdong Ma, Dianyin Cui & Lu Cao, "Genetic algorithm for the optimization of features and neural networks in ECG signals classification", Scientific Reports, January 2017.
19. Inan Gülera, Elif Derya Übeyli, "ECGbeat classifier designed by combined neural network model", Pattern Recognition, Elsevier Journal, pp 199-2008, June 2004.
20. Stanislaw Osowski, Linh Tran Hoai, Tomasz Markiewicz, "Support Vector Machine-Based Expert System for Reliable Heartbeat Recognition", IEEE Transactions on Biomedical Engineering, Vol 51, No 4, April 2004.
21. Jimena Rodríguez, Alfredo Goñi, Arantza Illarramendi, "Real-Time Classification of ECGs on a PDA", IEEE Transactions on Biomedical Engineering, Vol 9, No 1, March 2005.
22. Can Ye, Miguel Tavares Coimbra, B.V.K. Vijaya Kumar, "Arrhythmia Detection and Classification using Morphological and Dynamic Features of ECG Signals", 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, pp. 1918-1921, August 31 - September 4, 2010.

AUTHORS PROFILE



Mr. Sumanta Kuila is currently working as an Assistant Professor in the Department of Computer Science & Engineering, Haldia Institute of Technology, Haldia-721657, India. He has 16 years of experience in the field of Computer Science & Engineering. He is doing Ph.D in Computer Science & Engineering from Amity University Uttar Pradesh, Lucknow campus. He did his masters (M.E) from west Bengal University of Technology in 2005. He has done 6 Technical certifications (Sun & Oracle) on Java/J2EE & Web Technology. His current research interest includes Machine Learning, Data Analytics, Big Data, Internet of Things etc.



Dr. Namrata Dhandha is presently Professor in the Department of Computer Science and Engineering, Amity School of Engineering & Technology, Amity University Uttar Pradesh, Lucknow campus. She received her Ph.D from University of Petroleum and Energy Studies, Dehradun, India in 2015. She has 18 years of teaching experience in different reputed engineering & technical institutes in India. Her current research interest includes Data Mining, Big Data Analytics, Big Data, Data Wire housing Internet of Things etc.



Classification of Heart Arrhythmia in ECG Signals using PCA and SVM



Dr. Subhankar Joardar is presently Professor & Head in the Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia-721657, India. He received his Ph.D degree from Birla Institute of Technology, Mesra, Ranchi, India in 2016. He did his masters (M. Tech and MCA) both from BIT, Mesra, Ranchi in 2009 and 2002 respectively. He has published more than 14 technical papers in the referred journals/conferences. He has served as Organizing Chair of international conference (ICITAM 2017). He is also Program Committee member of International Conferences (ICCDC 2019). He is a member of Computer Society of India. His current research interests include Swarm intelligence, Routing in Mobile Ad Hoc Networks, Machine Learning.