Diagnosis of Congestive Heart Failure from HRV signal using SVM classifier and Patient Specific cross validation

Nabanita Sinha, Saurav Mandal

Abstract*: Congestive heart failure (CHF) is popularly known fatal cardiac disease that occurs when pumping action of heart is lower than normal case. The purpose of this study is to the accurate diagnosis of CHF by improving classifier performance with effective features extraction and cross validation approach. The identification of significant features in electrocardiogram is highly important to detect congestive heart failure. Therefore ,this paper introduces a classifier based automated detection scheme with a novel approach of feature extraction from Heart rate Variability (HRV) signal for early prediction of CHF. The dynamical characteristics of HRV signal is analysed by computation of Largest Lyapunov Exponent. The statistical features are also evaluated to capture crucial variation in HRV signal to distinguish the abnormal and normal heart condition. The extracted features are subjected to Support Vector machine (SVM) classifier for automated discrimination of CHF from normal ECG signal. Experimental results evaluate the performance of extracted features and estimate the accuracy of the classification using the 10 fold cross validation and patient specific cross validation approach. Our experiment is validated by ECG data of normal and CHF subjects from Physionet database. The proposed system is efficient to the detect CHF with an average accuracy of 98.75%,, sensitivity 98.38%, values and 98.94%. Based on comparative study with the existing scientific research work to diagnose CHF, our proposed approach is found to be reliable and efficient for CHF diagnosis

***Keywords* : HRV, Largest Lyapunov Exponent, SVM, ECG, CHF**

I. INTRODUCTION

Congestive heart failure (CHF) is a heart condition that results in the heart being unable to produce enough oxygen to various components of the body [1]. It is one of the cardiac disease which is leading cause of death according to the Reports of National Vital Statistics [2]. Its progression leads the wall of cardiac muscle to weaken so the heart's lower chambers are ineffective in pumping blood. CHF results in low blood flow that manifests fatigue and difficulty in daily operations. The main causes of CHF include high blood pressure, myocardial infarction, alcoholism and valvular disease [3]. Normally, electrocardiography (ECG) is used for detection of CHF. The variation in RR interval signal and the morphological variation in ECG signal are used to diagnose CHF [4]. For clinical practice, early detection of CHF is very important to minimize its progress.

Recent studies have proposed different techniques for automated identification of CHF by analysing ECG signal [5].Kuntamalla et al.[11] paper have applied sequential trend analysis with K-nearest neighbor classifier to find difference of CHF subjects from normal. Hossel et al. [12] established power spectral densities from the RR interval data and utilized to detect CHF from NSR. Yu et al. [15] have extracted feature with mutual information from heart rate variability features with Quadratic classifier to detect CHF. Isler and Kuntalap [16] had calculated feature related to entropy and used K-nearest neighbor to catagorised NSR and CHF. There are some proposed approaches where the features from the ECG signal have been used to recognize CHF symptoms. In the studies, of Masetic and Subasi[21] had applied the auto-regressive method on extracted ECG signal features. They used random forest classifier to detect the CHF patients. Sudarshan et al. [22] had employed the dual-tree complex wavelet transformed data using ECG signal features. The information had been given to the input of the classifiers like decision tree and KNN for classification of CHF and NSR. In this study we have introduced a strong approach that can analyze ECG signal more precisely .So that it can be useful to diagnose CHF with more accuracy. We have utilized the heart rate variability (HRV) signal for CHF detection. The ECG database is taken from MIT-BIH database. Various features are obtained from the HRV signal and fed to the Support Vector machine (SVM) classifier to differentiate Normal sinus rhythm of heart and CHF subjects. The chief impact of our proposed work is to evaluate the significant features from HRV signal and use both 10 fold cross validation and the patient specific cross validation of classifier performance for CHF detection.

II. ECG DATA SETS AND PROPOSED METHOD

Our proposed system for automated diagnosis of CHF 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. 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, the classifier is use for classification of CHF condition and normal sinus rhythm (NSR).

The extractions of significant features are very important to reduce computational complexity and increase detection accuracy of the

Revised Manuscript Received on January 5, 2020

***** Correspondence Author

Nabanita Sinha*, PhD Research Scholar, Razabazar Science College, University of Calcutta, Kolkata, India. Email:nabanitaroy.sinha@gmail.com

Saurav Mandal, PhD Research Scholar, Razabazar Science College, University of Calcutta, Kolkata, India. Email:souravsourav.cu@gmail.com

classifier. Thus, statistical analysis is performed to distinguish between normal and abnormal ECG data using machine learning algorithm. We have used different statistical features and nonlinear feature like largest lyapunov exponents. We have used first order statistical features-mean and standard deviation. The 2nd order statistical features skewness, kurtosis are also used in this study. All features extraction methods are described below.

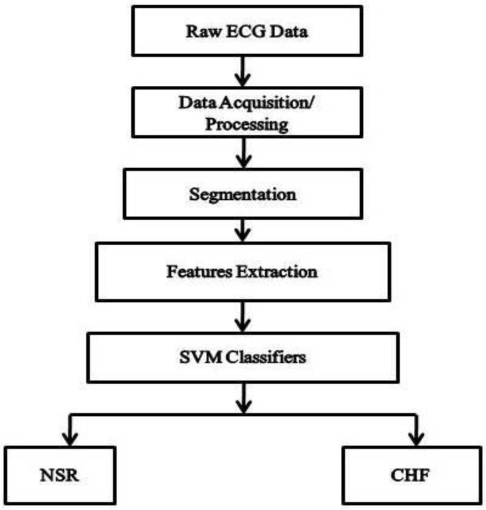

Fig. 1. Flowchart of Proposed System

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 ECG data signal from MIT-BIH databaseand each ECG are sampled at frequency of 128 Hz. The RR-interval database of CHF is used for our experiment. In this study, we have used 15 dataset of CHF patients and 30 dataset NSR subjects.

- ***Prepossessing***

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 0.3 Hz and notch filter of 50 Hz respectively[17].

- ***Segmentation***

After the pre-processing stage HRV signals are segmented for feature extraction and analysis. The HRV signals are taken of duration 12hr from physio.net database.The segmentation is done by taking input HRV signal where . each segment contains 600 samples. A total of 11,223 and 18,105 HRV segments are obtained from 15 CHF and 30 normal HRV time series data respectively.

B. Feature Extraction

- **Mean**

Electrical activity of heart signal is varying over time. Mean value of ECG signal is a noticeable feature we have found to distinguish between normal and abnormal subjects. Mean is simply average value of input data signals. It is represented by m. Mathematical representation of mean is

$$m = \frac{1}{N}\sum_{i=1}^{N} Z_i \quad (1)$$

where i=1, 2, 3,……, N and Z_i is input signal.

- **Standard deviation**

Standard deviation gives the amount of variation of data point respect to mean value. It is normally represented by the symbol σ. It is defined by square root of variance

$$\sigma(Z_1, Z_2, Z_3, \ldots, Z_N) = \sqrt{\mathrm{var}(Z_1, Z_2, Z_3, \ldots, Z_N)} \quad (2)$$

$\sigma = \sigma(Z_1, Z_2, Z_3, \ldots, Z_N)$ is the distribution of standard deviation.

Mean describe the location of distribution of time series data and variance give the result how far the data spread out from the mean.

$$\mathrm{Var}(Z_1, Z_2, Z_3, \ldots, Z_N) = \frac{1}{N-1}\sum_{i=1}^{N}(Z_i - \bar{Z})^2 \quad (3)$$

- **Skewness**

It is the measure the asymmetry of given distribution function around it mean. Positive value of Skewness indicates distribution with asymmetric tails move towards positive value and negative value indicates distribution with asymmetric trails move towards negative value. Skewness value zero indication of symmetric function. Tabachnick & B.G.Fdell [18] **s**hown that standard error of skewness is 6/N. N is represent sample number of data series. Distribution has significant skew if value of Skewness is more than 12/N. Means that data series is asymmetric in nature. Value of skewness with in significant range indicate that data series has no such skew problem.

$$\mathrm{Skew}(Z_1, Z_2, Z_3, \ldots, Z_N) = \frac{1}{N}\sum_{i=1}^{N}[\frac{Z_i - \bar{Z}}{\sigma}]^3 \quad (4)$$

- **Kurtosis**

Kurtosis is a statistical measure that shows relative peakedness or flatness of a distribution differs from normal distribution. Kurtosis value zero indicates normal distribution. Positive value of kurtosis indicates relatively peaked distribution and negative value is flat distribution. Mathematical representation of kurtosis is

$$\mathrm{kurt}(Z_1, Z_2, Z_3, \ldots, Z_N) = \frac{1}{N}\sum_{i=1}^{N}[\frac{Z_i - \bar{Z}}{\sigma}] - 3 \quad (5)$$

In this equation -3 term make the value zero of normal distribution.

- ***Largest Lyapunov Exponent***

Largest Lyapunov exponent (LLE) is an important invariant to detect chaos from a dynamic system. Dynamic system how much sensitive to initial condition is quantify by lyapunov exponent. For example, consider two trajectories of nearby initial condition on an attracting manifold. If the dynamic system is in chaotic in nature, then the trajectories is diverged at an exponential rate characterised by LLE [8]. Present of positive exponent is also sufficient for diagnosis chaos and

Retrieval Number: C7986019320/2020©BEIESP
DOI: 10.35940/ijitee.C7986.019320

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

representation of instability of dynamic system in a particular direction. Let Z(t) time evolution of initial condition Z(0) in a state space. Then LLE is found in state space

$$\lambda_1 = \lim_{t \to \alpha} \lim_{\varepsilon \to 0} \frac{1}{t} \ln\left(\frac{|Z(t)-Z_{\varepsilon}(t)|}{\varepsilon}\right), |Z(t)-Z_{\varepsilon}(t)| = \varepsilon \qquad (6)$$

Now we compute

$$P(\nabla t) = \frac{1}{N}\sum_{k=1}^{N} \ln\left(\frac{1}{\left\|\bigcup_k\right\|} \sum_{Z_t \in \bigcup_k} dist(Z_k, Z_t, \nabla t)\right) \qquad (7)$$

Where $\bigcup_k$ is the neighborhood of Z_k with diameter ε.

Intermediate range of ∇t. P(∇t) increase linearly with the slope λ_1.

III. CLASSIFIER

- **Support Vector Machine**

Classification and regression supervised learning method SVM is used [9,23]. It is belong the family to generalised linear classification. SVM has ability to minimize error and maximize the geometrical margin.

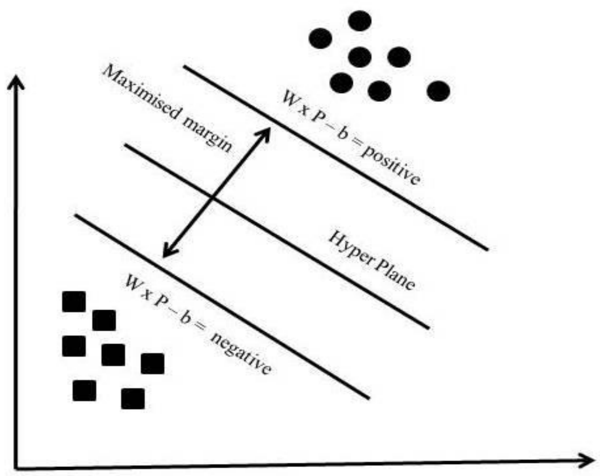

Fig. 2. Maximum margin of two hyperplanes for SVM and trained samples of two classes

Two parallel hyper plane are constructed both side of hyper plane. Distance between the two parallel hyper plane give better results of classification. We consider data points in P,Q plane as

$\{p_i, q_i\}_{i=1}^{N}$ with $p_i \in \{-1,1\}$ and $q_i \in R^n$, we use a C-SVM formulation []. To separate data by hyper plane using data minimization technique as given bellow

$$\tau(x,\sigma) = \frac{1}{2}\|x\|^2 + c\sum_{i=1}^{N}\sigma_i \qquad (8)$$

Where $p_i(x.q_i + b) \geq 1 - \sigma_i$, $\sigma_i \geq 0$ and $c \geq 0$ is the tradeoff in between constraint and regularization violation. In dual formulation we maximize in high dimensional feature space

$$x(\beta) = \sum_{i=1}^{N}\beta_i - \frac{1}{2}\sum_{ij}\beta_i\beta_j q_i q_j k(p_i p_j) \qquad (9)$$

where $0 \leq \beta_i \leq c$ and $\sum \beta_i q_i = 0$

Decision function is sign(h(p)) where

$$h(p) = \sum_{l=1}^{n}\beta_l q_l k(p, p_l) + b \qquad (10)$$

where $p_l : l \in \{1,2,3,\ldots.n\}$ will referred as support vector. Thus to classify a point n Carnel computation is needed in Kernelized SVM.

- **Cross Validation**

Cross Validation is used for the assessment of the predictive performance of the classifier. The motivation to perform cross validation techniques is to check classifier performance for different test data set. Without cross validation we have information only about our model performance to the given in-sample data. Cross validation is the re-sampling method which used for evaluation of classifier models on data having limited sample. In this study, the k-fold cross-validation method is used. The given data is spilt into k number of groups. We have chosen k=10 for our experiment. We have also applied patient specific cross validation approach to estimate the reliability of the proposed model. For this purpose, the HRV segments of particular CHF subject are taken and feed to the input of the SVM classifier for testing. The remaining ECG segments from other CHF subjects and NSR segments are taken for training of the classifier [10].

IV. RESULTS

In this work, the HRV segments are obtained from 15 CHF and 30 normal ECG recordings from MIT-BIH database.

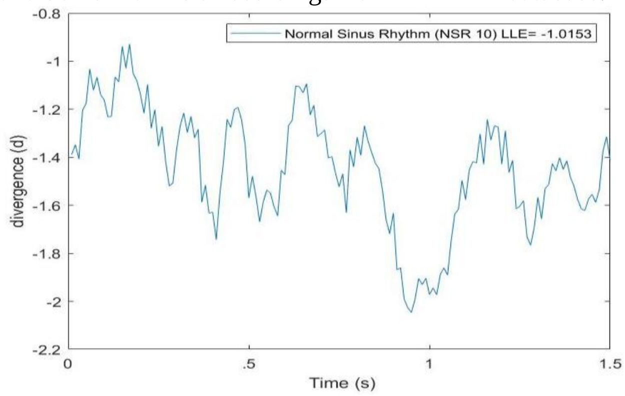

Fig. 3. Largest Lyapunov Exponent of Normal ECG signal(*NSR 10*)

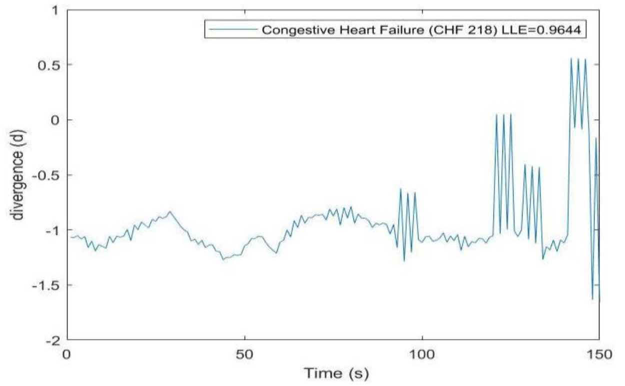

Fig. 4. Largest Lyapunov Exponent of Congestive heart failure patient ECG signal (*CHF 210*)

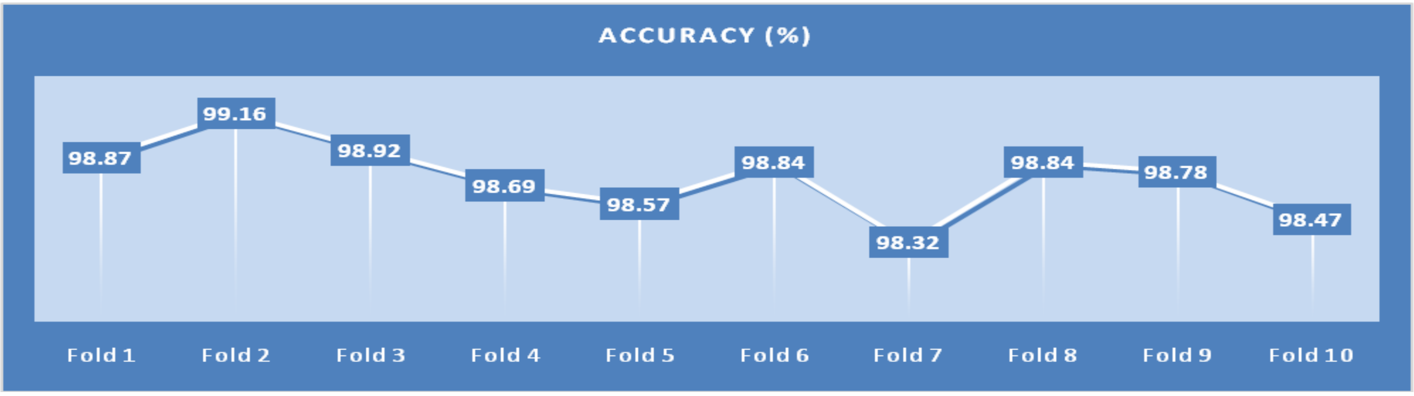

Fig. 5. Performce analysis of SVM classifier by 10-fold cross validation

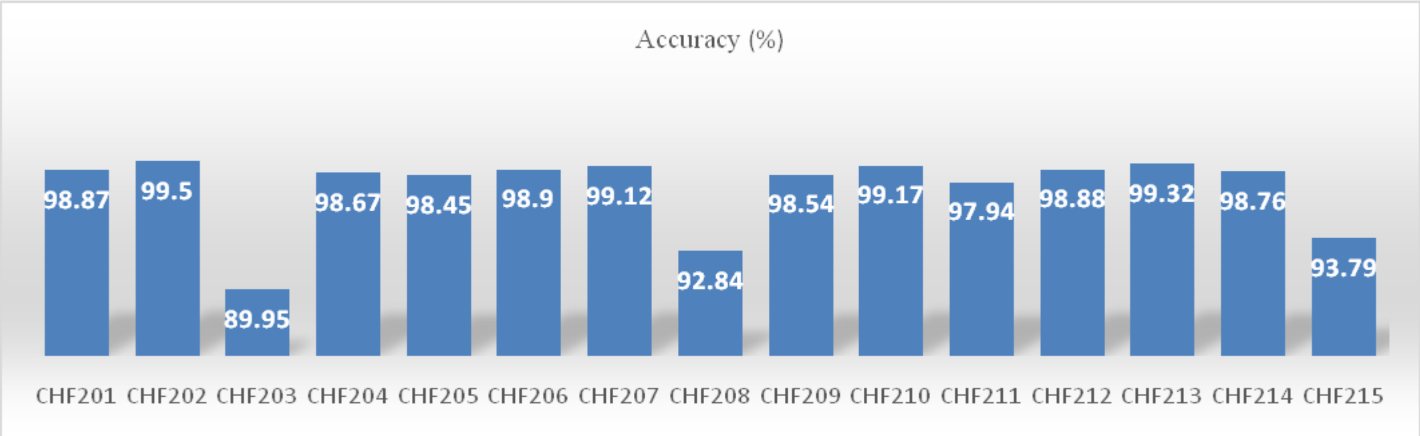

Fig. 6. Performance analysis of SVM classifier for individual patient.

The largest Lyapunov exponent is calculated from the HRV signal for both normal and CHF subjects as shown in Fig. 3 and Fig. 4 .The positive value of LLE confirms the presence of chaos in CHF signal .The dynamics of HRV signal is mostly non chaotic for normal subjects. The variation of LLE value for both normal and CHF is shown in Fig. 7. The estimated value of LLE exhibits clear separation between normal and CHF classes . The obtained values of statistic features from CHF and NSR HRV signal are shown in Table 1. The average performance of the classifier is evaluated by conventional indices such as specificity (Sp) , sensitivity (Sen), and accuracy (Acc) of the classifier as shown in Fig.8. Fig. 5 shows the performance of SVM classifier for 10-fold cross validation process. The patient specific cross validation is done by estimating the classifier accuracy for individual patients as shown in fig.6.

Table- I: Extracted Features Values for CHF and Normal ECG data.

Data	Skewness	Kurtosis	Mean	Standard Deviation
CHF	30 ± 15	1200 ± 200	0.66 ± 0.04	0.7± 0.3
Normal	0.8± 0.3	200 ± 25	0.74 ± 0.05	1.2 ± 0.04

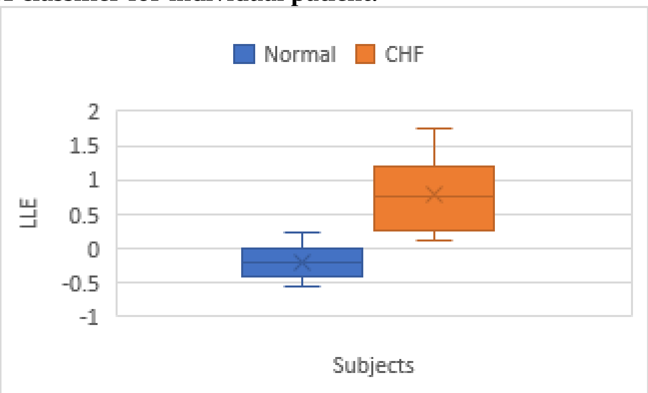

Fig. 7. LLE value for normal and CHF subjects

Fig. 8. Average Performance of SVM classifier

V. DISCUSSION

The main aim of the study is to diagnosis of CHF and normal heart condition by analyzing the HRV signal. The nonlinear dynamics of HRV signal is evaluated by measuring the mount of chaos present in normal and CHF ECG data. The fundamental and complex statistical features are evaluated to characterize the variation of HRV signal. It has been found that the extracted features provide clinical significance to classify normal and CHF subjects.

The value of LLE shows that chaos is present in CHF patients whereas the normal ECG data are non chaotic. From the obtained result, it has been found that the feature values for skewness and kurtosis are more significant to identify the disease comparing with other statistical features used in this work.

In this paper, we have introduced patient specific cross validation approach to evaluate the performance of our designed classifier. From the classification accuracy of individual CHF patients, it has been noticed that performance of SVM classifier is higher for most of the CHF patients. But the classifier performance for CHF203, CHF208, CHF 215 is degraded due to clinical variation in ECG signal. We have also achieved maximum accuracy of 99.5 % for CHF202. The obtained result from the patient specific cross validation approach shows our proposed model able to capture the changes in the above features measured for a patient undergoes having CHF with high accuracy.The overall performance of SVM classifier is estimated. The obtained result from our experiment and the existing research work for CHF detection are studied and compared. The comparison result is exhibited in table 2. The comparative analysis confirms that our proposed approach is quite efficient for CHF detection.

Table II: Comparison of proposed work with different reaserch work

	Year	Author	Classifier	Accuracy (%)
01	2010	Kuntamalla and Reddy [11]	K-nearest neighbour	96.68
02	2016	W. Chen[12]	SVM	96.67
03	2011	Pechia et al. [13]	Classification & Regression trees	96.40
04	2015	Kamath [14]	Threshold Classifier	98.20
05	2012	Yu and Lee [15]	Quadratic classifier	97.95
06	2019	A. Jovicet al [20]	RTF	90.7
07	2019	N. Saxena, L.S.Maurya [24]	SVM	90.23
08	-	Proposed System	SVM	98.75

VI. CONCLUSION

In this work we have proposed a model for computer aided diagnostic system to provide optimal treatment that can prevent the progression of CHF. Our proposed SVM classifier based CHF diagnostic model can distinguish normal and CHF subjects by analyzing HRV signal.The extracted nonlinear and statistical features from HRV signal are able to classify normal and CHF classes with high accuracy of 98.75%. Our proposed system is fully automated and reduce computational time for using a smaller number of features. The system is efficient to detect CHF within few seconds which save time for further investigation for severity of cardiac condition and treatment of CHF. In addition, the patient specific cross validation approach makes our proposed system more reliable and robust. In our future study we will focus on extracting new set of features extracted from more numbers of CHF patients to increase CHF detection accuracy of SVM classifier.

REFERENCES

1. G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. C. W. Yancy, M. Jessup, B. Bozkurt, J. Butler, D. E. Casey, M. H. Drazner, G. C. Fonarow, S. A. Geraci, T. Horwich, J. L.cardiology foundation/american heart association task force on practice guidelines," Journal of the American College of Cardiology,vol. 62, no. 16, pp. 1495–1539, 2013

3. M. Kenneth D. Kochanek, M. Jiaquan Xu, B. Sherry L. Murphy, M. Arialdi M. Minino, H.-C. Kung, Deaths: Final data for 2009, National VitalStatistics Reports 60 (3) (2011) 1–117.URL https://www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_03.pdf
4. S. Nadar, N. Prasad, R. S. Taylor, and G. Y. Lip, "Positive pressure ventilation in the management of acute and chronic cardiac failure: a systematic review and meta-analysis," International journal of cardiology, vol. 99, no. 2, pp. 171–185, 2005.
5. J. E. Madias, "The resting electrocardiogram in the management of patients with congestive heart failure: established applications and new insights," Pacing and clinical electrophysiology, vol. 30, no. 1, pp. 123–128, 2007.
6. G. Altan, Y. Kutlu, and N. Allahverdi ,"A new approach to early diagnosis of congestive heart failure disease by using hilbert–huang transform," Computer Methods and Programs in Biomedicine, vol-137, pp. 23–34, 2016.
7. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, *101*(23), e215– e220. Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218. doi: 10.1161/01.CIR.101.23.e215
8. J.P.Eckmann and D.Ruelle,"Foundation limitations for estimating dimensions and lyapunov exponents in dynamic systems",physica D56(185), 1992.
9. C. Cortes and V.Vapnik,"Support vector networks", Machine learning 20(3): 273-297,1995.
10. K. Daqrouq and A. Dobaie, "Wavelet based method for congestive heart failure recognition by three confirmation functions,"Computational and mathematical methods in medicine, vol. 2016, 2016.
11. S. Kuntamalla and L. R. G. Reddy, "Detecting congestive heart failure using heart rate sequential trend analysis plot," International Journal of Engineering Science and Technology, vol. 2, no. 12, pp. 7329–7334, 2010.
12. A. Hossen and B. Al-Ghunaimi, "Identification of patients with congestive heart failure by recognition of sub-bands spectral patterns,"2008
13. L. Pecchia, P. Melillo, M. Sansone, and M. Bracale, "Discrimination power of short-term heart rate variability measures for chfassessment," IEEE Transactions on Information Technology in biomedicine, vol. 15, no. 1, pp. 40–46, 2011
14. C. Kamath, "A new approach to detect congestive heart failure using detrended fluctuation analysis of electrocardiogram signals,"Journal of Engineering Science and Technology, vol. 10, no. 2, pp. 145–159, 2015.
15. S.-N. Yu and M.-Y. Lee, "Conditional mutual information-based feature selection for congestive heart failure recognition using heartrate variability," Computer methods and programs in biomedicine, vol. 108, no. 1, pp. 299–309, 2012.
16. Y. ˙I¸sler and M. Kuntalp, "Combining classical hrv indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure," Computers in Biology and Medicine, vol. 37, no. 10, pp. 1502–1510, 2007.
17. 17. Acharya, U. R., Faust, O., Sree, V., Swapna, G., Martis, R. J., Kardin. A., & Suri, J. S., "Linear and nonlinear analysis of normal and CAD-accepted heart rate signals", Computer Methods and Programs in Biomedicine, 2014, 113(1), 55–68
18. Tabachnick, B.G, Fidell, L.S ,"Using Multivariate statistics ", New York: Harper Collins, 1996.
19. Y Chung, HW Jung, J Kim, and JH Lee,"Personalized Expert-Based Recommender System: Training C-SVM for Personalized Expert Identification",Machine Learning and Data Mining,LNAI 7988, pp. 434–441, 2013.
20. A. Jovic, Brkic, K., &Krstacic, G. (2019). Detection of congestive heart failure from short-term heart rate variability segments using hybrid feature selection approach. Biomedical Signal Processing and Control, 53, 101583.
21. Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," Computer Methods and Programs in Biomedicine, vol. 130, pp. 54–64, 2016.
22. V. K. Sudarshan, U. R. Acharya, S. L. Oh, M. Adam, J. H. Tan, C. K. Chua, K. P. Chua, and R. San Tan, "Automated diagnosisof congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2s of ecg signals,"Computers in Biology and Medicine, vol. 83, pp. 48–58, 2017.
23. Sundari Tribhuvanam, H. C. Nagaraj, V.P.S. Naidu , "Arrhythmia Classification with Single Beat ECG Evaluation and Support Vector Machine", International Journal of Innovative Technology and Exploring Engineering (IJITEE) , Vol-8 ,Issue-12, October 2019
24. Niharika Saxena, L.S.Maurya , "Machine Learning Based Diagnosis and Prediction System for Congestive Heart Failure", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol-8 , Issue-10, August 2019

AUTHORS PROFILE

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.

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.